DOCUMENT RESUME

ED 237 085 IR 050 498

AUTHOR	Friberg, Ingegerd, Ed.
TITLE	The International Research Forum in Information Science Proceedings (4th, Boras, Sweden, September 14-16, 1981).
INSTITUTION	Boras Univ. Coll. (Sweden).
PUB DATE	82
NOTE	252p.
AVAILABLE FROM	Hogskolan i Boras, Bos 874, S50115, Boras, Sweden (50 Swedish Cr).
PUB TYPE	Collected Works - Conference Proceedings (021) -- Viewpoints (120) -- Reports - General (140)
EDRS PRICE	MF01/PC11 Plus Postage.
DESCRIPTORS	Action Research; Communication Problems; Databases; Documentation; Fiction; Information Dissemination; Information Needs; *Information Retrieval; *Information Science; Information Services; *Information Utilization; Language; Library Role; *Man Machine Systems; Models; *Online Systems; *Research Projects; Statistical Data; Use Studies
IDENTIFIERS	End Users; *User Needs

ABSTRACT

The Fourth International Research Forum in Information Science (IRFIS 4) focused on two themes: user-oriented research in information science and experimental and practical investigations of the mediation and use of information. An opening speech by Bjorn Tell reviewed information research policy in the United States, France, England, Japan, the Soviet Union, West Germany, and Sweden. Thirteen papers were presented, including: "Models of Dialogue for Information Retrieval," by Nicholas J. Belkin; "Language as a Barrier to Communication," by Ralph Adam; "Action Research and Users' Needs," by T. D. Wilson and D. R. Streatfield; "Education of the End-User in Computerized Information Retrieval," by Nancy Fjallbrant; "The Dissemination and Use of Scientific and Professional Information in Medicine," by Tuula H. Laaksovirta; "Project Athena, 'Thinking Aloud,' and User-Oriented Research," by Kathleen T. Bivins; "The Concept of Entailment for the Purpose of Mediation," by Marta Grabowska; "Factual Data Retrieval from a Data Base Constructed According to the SAP-Technique: A Project Incorporating Government Committee Reports," by Bjorn V. Tell and Irene Wormell; "The Wasteful Use of Empirical Statistical Information in Information Research," by Bertram C. Brookes; "Research on the Needs and Uses of Scientific and Technical Information: Research Areas and Problems," by Lars Hoglund and Olle Persson; "The Librarian's Role as a Mediator in Fictive Literature," by Annelise Mark Pejtersen; "Problem Solving and Information Retrieval: A Case," by Povl Timmerman; and "Pitfalls of User Oriented Research, and Some Neglected Areas," by J. M. Brittain. Also included in the conference proceedings are workshop reports on models and theories of users' needs, influences on information science from other disciplines, and practical research applications and projects. A conference schedule and a list of participants are provided. (ESR)

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED237085

THE 4TH INTERNATIONAL RESEARCH FORUM IN INFORMATION SCIENCE

Proceedings edited by

Ingegerd Friberg

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
Kerstin Quentzer

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

IR050498

SKRIFTER FRÅN HÖGSKOLAN I BORÅS
BIBLIOTEKS- OCH INFORMATIONSVETENSKAPLIGA STUDIER 1

ISSN 0280-381X Skrifter från Högskolan i Borås
ISSN 0280-3828 Biblioteks- och informationsvetenskapliga studier

Minab/Gotab 1982

FOREWORD

The Fourth International Research Forum in Information Science was held in Borås, Sweden, during September 14-16, 1981. The previous conferences were held in London in 1975, in Copenhagen in 1977, and in Oslo in 1979.

The organization of this forum was made possible by financial support from The Delegation for Scientific and Technical Information and from The University College of Borås.

Two themes were set up for the conference: "User-Oriented Research in Information Science" and "Experimental and Practical Investigations of the Mediation and Use of Information". Papers on both themes were presented at the conference and aspects of both themes were dealt with in the workshops.

In editing the proceedings I have not tried to group the papers around these themes. Instead, I have chosen to follow the actual order of the programme, in the hope of making it possible for the participants to reexperience the presentations as they were given, and, what is more important, showing how theory and practice are integrated and should not be isolated from each other. The reports from the workshops form the conclusion to the proceedings, with the report from Workshop 3 giving an overview of the whole conference in its commentaries on each paper in order of presentation.

Finally, I want to extend my thanks to those who have entrusted me with the task of editing the proceedings, and to Bertil and Kerstin Quentzer for giving me their support during the work, and last but not least to Carina Waldén for typing the manuscripts.

Gothenburg, February 25, 1982 Ingegerd Friberg

TABLE OF CONTENTS

CONFERENCE PROGRAMME 1

OPENING OF IRFIS 4 4
Welcoming address by Nils-Bertil Faxén, University College of Borås

INTRODUCTION TO IRFIS 4 5
Opening speech by Björn Tell, University of Lund

MODELS OF DIALOGUE FOR INFORMATION RETRIEVAL 15
Nicholas J. Belkin, Centre for Information Science, The City University, London

LANGUAGE AS A BARRIER TO COMMUNICATION 37
Ralph Adam, The City University, London

ACTION RESEARCH AND USERS' NEEDS 51
T.D. Wilson, Department of Information Studies, University of Sheffield, and
D.R. Streatfield, National Foundation for Educational Research, Slough

EDUCATION OF THE END-USER IN COMPUTERIZED INFORMATION RETRIEVAL 71
Nancy Fjällbrant, Chalmers University of Technology, Gothenburg

THE DISSEMINATION AND USE OF SCIENTIFIC AND PROFESSIONAL INFORMATION IN MEDICINE 91
Tuula H. Laaksovirta, The Academy of Finland, Helsinki

PROJECT ATHENA, "THINKING ALOUD", AND USER-ORIENTED RESEARCH 114
Kathleen T. Bivins, SCANNET, Gothenburg

THE CONCEPT OF ENTAILMENT FOR THE PURPOSE OF MEDIATION ... 127
Marta Grabowska, Warsaw University

FACTUAL DATA RETRIEVAL FROM A DATA BASE CONSTRUCTED ACCORDING TO THE SAP-TECHNIQUE. - A PROJECT INCORPORATING GOVERNMENT COMMITTEE REPORTS 135
Björn V. Tell & Irene Wormell, University of Lund

THE WASTEFUL USE OF EMPIRICAL STATISTICAL INFORMATION IN INFORMATION RESEARCH 150
Bertram C. Brookes, London

RESEARCH ON THE NEEDS AND USES OF SCIENTIFIC AND TECHNICAL INFORMATION. RESEARCH AREAS AND PROBLEMS 158
Lars Höglund & Olle Persson, Inforsk, Department of Sociology, University of Umeå

THE LIBRARIAN'S ROLE AS A MEDIATOR IN FICTIVE LITERATURE . 178
Annelise Mark Pejtersen, Royal School of Librarianship, Copenhagen

PROBLEM SOLVING AND INFORMATION RETRIEVAL. A CASE 208
Povl Timmerman, Royal School of Librarianship, Copenhagen

PITFALLS OF USER ORIENTED RESEARCH, AND SOME NEGLECTED AREAS 213
J.M. Brittain, Department of Library and Information Studies, Loughborough University

REPORT ON WORKSHOP 1 : MODELS AND THEORIES OF USERS' NEEDS 228

REPORT ON WORKSHOP 2 : INFLUENCES FROM OTHER DISCIPLINES: BORROWING AND LOANING 233

REPORT ON WORKSHOP 3 : PRACTICAL RESEARCH APPLICATIONS AND PROJECTS 237

LIST OF PARTICIPANTS 242

IRFIS 4 - CONFERENCE PROGRAMME

September 14 - 16, 1981

Monday, September 14

09.00-10.00	Registration and Coffee	
10.00-10.30	Opening of Conference Nils-Bertil Faxén and Björn Tell	
10.30-12.00	N J Belkin,	Models of Discourse and Information Retrieval System Design
	Ralph Adam,	Language as a Barrier to Communication in the Social Sciences
12.00-13.00	Lunch	
13.00-15.00	David Streatfield,	Action Research and Users' Needs
	Nancy Fjällbrant,	The Education of the End-User in Computerized Information Retrieval
	Tuula H Laaksovirta,	The Mediation and Use of Medical Information
15.00-15.30	Coffee	
15.30-17.30	Workshop 1: Models and Theories of Users' Needs	
	Introduction by Kathleen Bivins	
	Workshop 2: Influences from Other Disciplines: Borrowing and Loaning	
	Introduction by Povl Timmerman	
	Workshop 3: Practical Research Applications and Projects	
	Introduction by Tor Henriksen	
19.00	Conference Dinner	

Tuesday, September 15

08.30-10.00	Kathleen Bivins,	Methods in User-Oriented Research: A Critical Review of "Thinking Aloud"
	Marta M Grabowska,	The Concept of Entailment for the Purpose of Mediation
10.00-10.30	Coffee	
10.30-11.30	Björn Tell and Irene Wormell,	Factual Data Retrieval from a Data Base Constructed according to the SAP-Technique. - A Project Incorporating Government Committee Reports (Incl a demonstration of a data base)
12.00-13.00	Lunch	
13.00-14.30	B C Brookes,	The Wasteful Use of Empirical Statistical Information in Information Research
	Lars Höglund and Olle Persson,	Research on the Needs and Uses of Scientific and Technical Information - A Programme for Research in Sweden
14.30-15.30	Workshop 1: Models and Theories of Users' Needs Chairperson: Kathleen Bivins Workshop 2: Influences from Other Disciplines: Borrowing and Loaning Chairperson: Povl Timmerman Workshop 3: Practical Research Applications and Projects Chairperson: Tor Henriksen	
15.30	Dinner	
17.00	Departure by bus for a visit to Stora Teatern, Gothenburg, including a performance of the opera Mahagonny	

Wednesday, September 16

08.15-09.45	Annelise Mark Pejtersen,	Theory and Practice of the Librarian's Role as a Mediator in Fictive Literature
	Povl Timmerman,	Problem Solving in Information Retrieval: How can Library Practice Gain from Cognitive Psychology, Linguistics, Social Science, Computer Science, and Educational Psychology?
09.45-10.15	Coffee	
10.15-11.00	Michael Brittain,	Pitfalls of User-Oriented Research and Some Neglected Areas of Research
11.00-12.00	Workshop 1: Models and Theories of Users' Needs Chairperson: Kathleen Bivins Workshop 2: Influences from Other Disciplines: Borrowing and Loaning Chairperson: Povl Timmerman Workshop 3: Practical Research Applications and Projects Chairperson: Tor Henriksen	
12.00-14.00	Lunch sponsored by the city of Borås	
14.00-15.30	Summing up of IRFIS 4: Report from Workshops Planning for IRFIS 5	
15.30	Farewell Coffee	

OPENING OF IRFIS 4

Welcoming address by Nils-Bertil Faxén

On behalf of the University College of Borås I would like to welcome all of you, participants in the 4th International Research Forum in Information Science, to Sweden, to the University College, and to the Swedish School of Library and Information Science.

This is the first international conference that has been held at this college, and the city of Borås is, of course, very proud to follow in the lead of such cities as London, Copenhagen, and Oslo, where the earlier IRFIS-conferences have taken place. To many of the Swedish participants it might seem a little odd that a research forum is being held at a college where there are no research departments, no postgraduate study programmes, and no professors. This fact is due to the organization of higher education in Sweden today and to the fact that so far the library and information science has not been formally established in the higher educational system. But this does not mean that there is a lack of interest in information science and research, although no papers will be presented during the conference by our own people. We all hope that this will be changed before IRFIS 5 takes place.

The time allotted for planning the conference has been short, but I know that those responsible for the arrangements, Dr Ingegerd Friberg and Lecturer Bertil Quentzer, have done their utmost to give the conference a good setting. I hope that it will be successful, and that you will return home stimulated by the interesting reports that are going to be presented, having found new colleages and friends, revived earlier friendships, and placed the University College and the city of Borås on your own map of Europe.

INTRODUCTION TO IRFIS 4

Opening speech by Björn Tell

In the history of the International Research Forum of Information Science, Sweden has been a latecomer. Three fora have been successfully completed with prominent contributions not only from countries such as Great Britain or West Germany, but more so from our Scandinavian counterparts Denmark and Norway, but hitherto without Swedish contribution. The proceedings resulting from the earlier fora have been published and reached a wide audience. Thus, the International Research Forum on Information Science can be said to have become an institution.

To preserve its vitality there must be continuous efforts to arrange new fora on a regular basis, and through the proceedings assure that the research results get a wide dissemination. At the Oslo meeting the representative of the Federal Republic of Germany extended an invitation to IRFIS 4 to be held in 1981. Now, we are writing 1981 and you are in Sweden. That came about because late in 1980 the Gesellschaft für Information und Dokumentation, through its Director, Professor Werner Kunz, expressed their regrets that due to administrative and organizational circumstances the Gesellschaft was not able to undertake the organization of IRFIS 4 as expected. Instead, they would appreciate it if Sweden undertook that task.

As time had passed, that put us in a rather difficult position. The time for planning became very short. Those of you who have submitted papers know, I presume, that the call for papers we sent out did not give you the necessary time to spark off new research ideas or complete work in progress in such a way that the reporting could live up to the expected standards of IRFIS.

On our side we had to find a suitable conference place, a secretariat and funds for carrying through the arrangements. After some trial and error, the Swedish School of Library and Information Science at the University College of Borås accepted the responsibility to arrange the forum. A secretary able to devote time to the organizational work was found, and most important of all the National Delegation for Scientific and Technical Information gave the needed financial support, even for the publication of the IRFIS 4 proceedings.

As hosts now, we are most happy to extend a hearty welcome to all of you at this opening of the Fourth International Research Forum in Information Science. We are glad that the late call for papers did meet with understanding, as we have received 15 papers so far. Many of the contributors have been represented at earlier fora. From the side of the organizers we thus assume that IRFIS has gone through a shake-down process, so that the presentation of papers will spark off stimulating discussions. The short preparation time could serve as a time at this forum for summing up, to reflect upon earlier achievements, and to form a basis for new initiatives to be presented at future fora. Dr. Kunz has now in writing confirmed that the DGD Sektion für Informationswissenschaften will take over the responsibility for the planning and organization of IRFIS 5 in 1983 in the Federal Republic of Germany.

Information research and research policy

On this occasion it would be tempting to look back and try to identify those research areas which have been the object for study by those who have attended the earlier IRFIS conferences. However, many of you may be more familiar with that than I am, because as I mentioned, Sweden is a latecomer in the gatherings of IRFIS, although we have followed the activities through the proceedings. So instead, I will try to review the research policy that has been developed over the years in some countries at the leading edge.

Does the United States have a research policy?

The United States is regarded as a forerunner in the field of information science. Both that concept and IRFIS are created out of American predecessors. However, it may be shocking to hear that the United States does not have a ubiquitous information research policy. Every federal agency decides upon its research input in the information field, and sets its policy according to the funds available within its budget. Notable research is carried out in many federal agencies. For instance, the National Library of Medicine does research in the applications field using plasma displays and video-disc techniques linked to minicomputers. The NASA remote sensing activities have had a spin-off for storage of images and full text in machine readable form for retrieval and transfer by satellite communications. Only the Department of Energy is spending around 30 M dollars on its information and documentation activities, and that gives it plenty of latitudes for research. Generally, we can say that the situation in the United States is fragmented, especially, if all the work going on in the universities and in the private industry is taken into account. However, fortunately our American friends have excellent tertiary publications that allow us to follow their achievements.

The White-House Conference in 1979 did not result so much in the establishment of an information research policy, as more in the creation of an understanding of the need to increase the budget for the National Science Foundation so that it can devote more efforts to the field. However, because of the change of the administration, the NSF may not have succeeded in getting a rise for information work, and the present level is far below what was available during the 60's.

That does not mean that the NSF lacks a research policy in the information field. Last year it devoted around 5 M dollars to four areas, namely

1. Standards and measures for information science, i.e. a program concerned with the definition of objectives and quantitative measures of attributes of information, such as quantity, complexity, meaning, utility, and value, and their relation to the structural characteristics of information and its organization.

2. The structure of information, i.e. the study of information as idealized organization or structure from the standpoint of natural science.

3. Behavioral aspects of information transfer, i.e. the research of the abilities and limitations of people as information processors.

4. Infometric models, i.e. the interactions with the rapidly changing microelectronic and telecommunications developments call for a more comprehensive and fundamental understanding of the role of information as a component of the economy.

As part of the policy, NSF has instituted a 'Research Investigation Award' in order to attract competent researchers to the field. In 1979, more than a dozen such awards were given.

France, strong in applications research

For years the BNIST - Bureau national de l'information scientifique et technique - has been performing the French research policy. In 1979 it was transformed into MIDIST - Mission interministerielle de l'information scientifice et technique. MIDIST had as a sort of 'seed money' a budget in 1979 of 2,58 M Frs that through a multiplicator effect generated research for around 10 M Frs. While BNIST reported to the Minister of Industry the MIDIST reports directly to the French Prime Minister.

A decisive factor for research in the information area is that the policy should be related to the research policy in general, especially, the various applications that the policy gives rise to. The American examples have greatly influenced the French policy to broaden and strengthen its responsibilities for research and development in the information field. Thus, most of the funds have been spent in order to establish data bases, the building up of networks, and to promote the training of information specialists.

As a result of that policy around fifty new data bases are coming into being, a few examples outside the ordinary librarian's field can be mentioned, namely the High Pressure Data Bank (Prof. Vodar) and the Thermodata and Pluridata Data Banks (Profs. Bonnier & Dubois), both national achievements of international significance. In the network area the establishment of TRANSPAC has made it possible to spread the usage of the new data bases now available in France. Let us hope that France one day discovers IRFIS and wants to collaborate even though the working language will remain English.

British research policy far out in the applications field

Since the late 50's Great Britain has been determined to pursue an information research policy. The foundation of the British Library that took over the OSTI policy for research, has been instrumental for R&D activities on the British Isles. Complementary to the more basic research policy that has been expressed by the National Science Foundation in Washington, the British Library has been lax in its apprehension of what is meant by basic research, and has let it go far out into the applications field.

Like the French the British have thought it important to give heavy support to applications when introducing the new information technology. In 1979 the research budget of the British Library was around £ 1 M. Part of that has been spent upon the purchase of hardware. Thus a number of Hewlett Paccard minis have been acquired and given out to various research projects. The British Library also attaches importance to user studies that can be witnessed by the earlier contributions to IRFIS. Support has also been given to the classification field, e.g. to PRECIS even though that support may dry up. Another support to applications is that to computerized services such as BLAISE.

Japan towards the information society

Japan's research policy has been formulated by the Ministry of Education, Science and Culture. Support is given to basic research in the information field carried out in various universities. It has also taken the initiative for the planning of the future of integrated information activities by publishing 'A new plan of science information systems in Japan - An interim report, June 1979'.

It should be noted that Japan like Sweden does not yet have an academic chair in information science. Instead, such studies are integrated in a package together with information technology and computer science. Earlier difficulties to transfer Chinese and Japanese characters into machine-readable form seem now to have been resolved. A 20x24 dot matrix has been agreed upon, and over 10.000 characters have now got a standardized form. That means that the policy will aim primarily to build up Japanese data bases.

Long range policy, the mark of Soviet policy

The Soviet information policy is part of a 20-year plan that is based upon forecasts which continuously are revised in order to keep pace with changes in the environment. Also in Soviet the information policy should serve and support the R&D activities in general. Methods for scientific and technical information transfer are given high priority.

In general, the research work is directed towards mathematical and theoretical models for information transfer. In the Soviet Union like in Japan the linguistic problems are of importance. There is a need for solutions in order to be able to establish networks between the various republics and their languages and alphabets. A proper infrastructure is needed because of the geographical differences. A number of scientific committees with a broad representation from universities and state agencies have been established in order to define new research objectives, and to assess on-going projects.

Policy in the host country of IRFIS 5

As we are meeting next time in the Federal Republic of Germany, we then will be more acquainted with its research policy. However, let me mention the following. The invitation for IRFIS 5 is extended by the Gesellschaft für Information und Dokumentation - GID. That body was established in 1977 by integrating Das Institut für Dokumentationswesen, Die Zentralstelle für maschinelle Dokumentation, and Die Studiengruppe für Systemforschung. The present budget is around 16 M DM, and there are around 250 persons on the staff. GID's information policy is also broad and a few priority areas can be mentioned, such as

- Promotion of information and documentation for research and development
- Development of an information research policy after a study of objectives
- Documentation languages and classification systems, particularly the methods of construction, and the relationship between structure and the subject terminology
- User research studies, particularly problems of user analysis and user behavior, needs for information, elaboration of criteria for demand analysis in the design of information systems
- Research in training for information and documentation
- Planning and design of prototypes of information systems, and research in information and documentation technique vs. existant hardware
- Research and development in specialized fields, e.g. chemical documentation

The GID's funding activity goes into the field of publication and even bookselling and dissemination, but does not embrace mass media activities.

Research policy - the Swedish way

The Swedish information policy that eventually crystallized may be different from that of the big countries. For a small country it has seemed necessary to promote research and development activities that could facilitate the acquisition, processing and dissemination of scientific and technical information in view of the new computer and communications technology that opened up interesting potentials for the efficient transfer of information.

The policy has been to acquire a number of data bases in machine readable form and process them for SDI purposes. Thus, the MEDLARS operations of Karolinska Institutet and the INSPEC and fifteen other data bases running at the Royal Institute of Technology were an outcome of development and implementation work during the first decade of SINFDOK funding.

In 1973 support was primarily given to the development of a system for on-line information searches, and in parallel funding from other sources supported the development of LIBRIS - the cataloging and localization system for materials acquired in the research libraries. The funding was taken over in 1979 by the Delegation for Scientific and Technical Information that also is sponsoring this forum. It has a budget around 18 M Sw Kr equal to the same amount in French Francs.

The aim of the new delegation is to support in equal parts R&D-activities in the information and the library field. The policy has aimed to increase the utilization of information in data banks. Sweden then prompted the other Nordic countries to set up a packet-switched network, SCANNET, that now has been operating since 1976, and I can add, with much higher efficiency than the EURONET.

The success in the network field has furthered the incorporation of fact-retrieval data bases as complements to the bibliographic ones. Thus, the Cambridge Crystallographic Data Base and the NBS Mass Spectrometry Data Base have been implemented for searches through SCANNET. Sweden's participation in international projects such as INIS, AGRIS, MEDLARS, IRRD has since long taken the form of machine readable products. As an example of the establishment of a national data base the Government Committee Reports Data Base will be the subject for one contribution to this IRFIS.

Considerable efforts have been devoted to user studies, but the results may seem meagre. Instead, more progress has been made in the field of training. An interesting development in training will be reported from the Chalmers Institute of Technology.

The research policy also aims to detect future potentials. The use of video technique for storing both bibliographical texts and images in a case of the description of museum artefacts is under study. So is also the possibilities to transfer full texts by satellite through the work in the European Space Agency. In order to draw younger people into the field a fellowship program is in vigor.

Research policy for the future

From the overview I have given, it can be seen that many of the Western countries have established special funding agencies in order to promote and administer R&D activities in the information field. The ISORID program of Unesco, which for 1979-80 has a budget of 117.000 dollars, is supposed to cover the need for information about who is doing what in research in information. However, in spite of the close collaboration in this endeavour with FID, IFLA and ICA, many countries do not report, and essential works in the big countries are not always found in the ISORID data base. That means that the funding agencies, relying upon such an intergovernmental resource as ISORID, will have difficulties to assess research applications and projects which are not directly linked to specific R&D activities known from other places.

For younger researchers there should be more fellowship programs for international exchanges. The present fellowship program of Unesco primarily has as its aim to transfer trainees from the developing countries to places where they can be trained in the daily routines of library and information work. What is needed is also a fellowship program for young researchers in the field to go to centers of excellence and continue research.

In many countries there is a shortage of knowledgeable people who can take advantage of the new technology to the benefit of their country. The present transparency in the information field of all Western countries creates great potentials for their work as information specialists. But the profession needs also information scientists. Let us hope, therefore, that the funding agencies give more support and promote information research projects, so that more young people become attracted to the field.

The initiation of IRFIS and its persuance should in that context be seen as a catalyst not only to spread the knowledge about the continuous need for information research, but also to confront those already in the field with the research standards and the international research frontier, so that badly conceived projects or duplications of work can be avoided.

It seems essential that IRFIS becomes vigorous so that it will attract more researchers from still more countries into its framework. As I mentioned in the beginning, there is no hope that IRFIS 4 will be a milestone in the IRFIS history, because of the short time for preparation. However, during the sessions here, we will have time for afterthoughts, so that new research activities carrying innovative ideas will materialize for future conferences.

MODELS OF DIALOGUE FOR INFORMATION RETRIEVAL

Nicholas J. Belkin

Introduction

Today, I feel a bit like I have come full circle, at least as far as the sequence of IRFIS meetings is concerned. This is because I have no empirical evidence, no data, no experimentation to report here, but rather a pre-theoretical construct, which may lead to a research program. This is a bit like going back to my presentation, with Stephen Robertson, at the first Forum (Belkin & Robertson, 1976), which I now see as the beginning of a research program that was subsequently reported at the Copenhagen (Belkin, 1980a) and Oslo (Belkin, Brooks & Oddy, 1979) meetings, each presentation becoming successively more concrete and empirical (at least in my optimistic estimation). At the last meeting, we finally presented results of an experimental research project which embodied some of the ideas presented in the previous two Fora: today I would like to take a step back, and try to set up a framework which might subsequently result in further research projects. This presentation is, however, informed by our previous work, and I hope represents an expanding spiral rather than a full circle in terms of progress.

The problem which I would like to discuss is the nature of interaction between human and machine in machine-based information retrieval systems. In particular, I am concerned with information retrieval systems which are designed to be used by humans who are not necessarily experts with the system, who do not necessarily have well-specified needs and who have no great desire to learn a conventional means for communicating with the system. I am furthermore concerned with, in the general case, systems whose clients are unpredictable in terms of characteristics such as subject of interest, degree of expertise in the subject, type of information need or problem, but who are similar at least in their general cultural context if not in their preferred language of communication. In short, the general model is of a large scale public information system, such as people have envisaged arising with the advent of viewdata systems.

About such information retrieval systems, and indeed information retrieval systems in general, I have some fairly strong assumptions. These have been specified to some extent in our publications on Anomalous State of Knowledge Based Information Retrieval, especially Belkin (1980b), and include, importantly in this context, the necessity of interaction between user and IR mechanism, and the undesirability of requiring the user to specify her/his 'information need'. On at least the first of these points, we are by no means alone, as many people, in information science and outside it, have suggested that the interface between human and machine in human-machine systems is of primary importance in their effectively achieving their goals, and that such an interface may have to be an interactive one. Our work (especially including Oddy's, 1975, THOMAS) seems to me, by its emphasis on non-specifiability of need, to make such interaction absolutely necessary. Therefore, I feel it is incumbent upon us, at least, to specify in some way how such interaction should, or ought to take place. What I would like to do today is to outline a framework within which work at such specification could take place. My hope is that this framework will be general enough so that it will not be limited in its application to just ASK-based systems, but I do nevertheless assume that it will be difficult to apply it straightforwardly to conventional large-scale IR systems (except via intermediary devices).

Interaction in information retrieval and the IR situation

I would like to make my notion of the type of interaction appropriate to information retrieval more specific. This entails my going back to previous work, and introducing some concepts and figures that must be overfamiliar to at least some of you. My apologies.

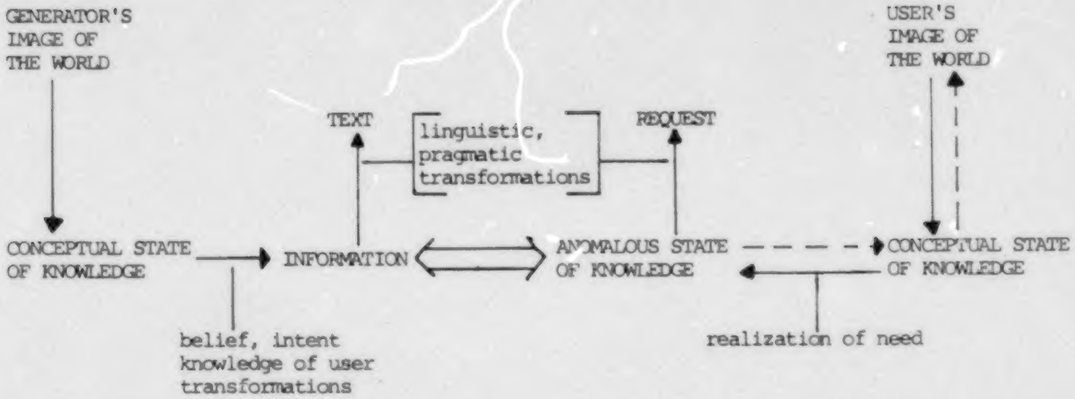

GENERATOR'S IMAGE OF THE WORLD
CONCEPTUAL STATE OF KNOWLEDGE
belief, intent knowledge of user transformations
INFORMATION
TEXT
linguistic, pragmatic transformations
REQUEST
ANOMALOUS STATE OF KNOWLEDGE
USER'S IMAGE OF THE WORLD
CONCEPTUAL STATE OF KNOWLEDGE
realization of need

Figure 1. A cognitive communication system for information retrieval.

Figure 1 manages to encapsulate many of my ideas about information retrieval. In particular, it specifies the IR situation as a recipient-instigated and recipient-controlled communication system, meaning that it is the user's problem which is paramount and which has instigated the communication exchange, and that the success of the exchange ought to be judged in terms of the user's satisfaction with its outcome. This often, but certainly not always, will be judged in terms of resolution of the anomalies in the user's state of knowledge, or even more basically, in terms of resolution of the user's problem.

Another important aspect of this diagram is the ASK: this is explicit recognition of the idea that the user is unable to specify (in the general case) what changes in the state of knowledge will be appropriate for the solution of the problem. What I would like to stress here is the dynamic nature of the system; as information is acquired, state of knowledge changes, perception of problem changes, perception of need changes, image of the world changes, and therefore anomalies change. This implies, at least to me, that IR is basically an iterative situation, and that it is interactive, in a specific sense.

This interaction is indicated in Figure 1 at the double arrow between user and text. In this diagram, a great deal is subsumed by this symbol, and it might perhaps be useful to make some of it explicit, as have, for instance, Ingwersen and Kaae (1979). This is the point of contact between user and mechanism, and it can be variously interpreted as reference librarian, intermediary, search strategy, matching algorithm, and so on. Its essential character is that it provides the means by which the mechanism 'understands' the user's ASK (or problem) so that it (the mechanism) can make a guess as to which text(s) in the corpus might help the user in resolution of ASK (or solution of problem). Since the system is dynamic, as indicated before, the understanding of ASK will change with each iteration. At this level, interaction will consist of descriptions by the user of the user's (changing) ASK, so that the mechanism's model of ASK, and perhaps of texts, can be modified appropriately. Of course, there will be other aspects to the interaction as well. I would like to make a terminology shift, and now call what I have referred to as the 'mechanism' the intermediary. This is intermediary in the most general sense, however, of that which lies between user and corpus.

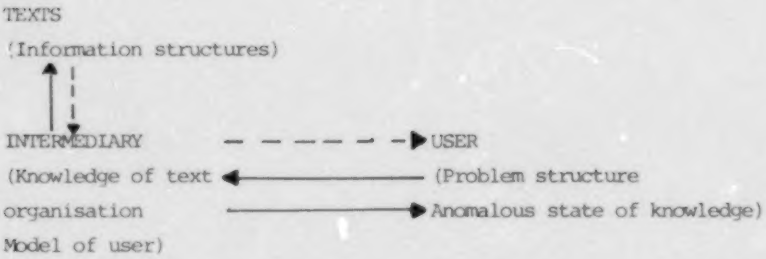


Figure 2. Some interactions among user, intermediary and texts.

If we consider the double arrow separately, then, we discover a communication system which two participants, mechanism and user, who have a common goal, resolution of user's ASK (or better, eventual solution of user's problem) and who intend to achieve this goal by interaction with one another and by interaction with a corpus of texts. Notice especially that the user's access to texts is only via the intermediary.

This, then, is the context of the interaction which I see occurring in the information retrieval environment. It is two-party, with an external data base to which both parties have recourse. It is goal oriented, with the goals of one of the parties (the user) being primary, and it assumes (or hopes) that the data base will be able to help the user to attain the goal. The two parties each have special, distinct roles to play in attempting to attain the goal, and they engage in their interaction with one another and with the data base until the user feels that the goal has been attained or decides to stop the interaction for some other reason. I would like to restrict the general domain of this interaction for our purposes here by specifying that the non-user party is a machine; that is, that we consider the case in which the user interacts directly with the intermediary which also manipulates the corpus, without a human intermediary. Furthermore I wish to consider the case in which the user is not necessarily a frequent, expert user of the system. These restrictions, I believe, are valid for the context of large-scale public information systems, at least in the short-term future.

It will be noticed, I hope, that the interaction as specified has some of the characteristics of normal conversation, and of certain types of restricted dialogue, such as interviews. I would like to suggest now that the interaction between human and machine in this type of information retrieval system will be most effective if it is treated as a conversation, and that the systems ought to be designed as conversations between mechanism and user. I make this suggestion because it appears to me that the context of information retrieval systems as specified above requires that interaction be as familiar as possible to the user, in order to make discovery and representation of need or problem as effective as possible. Modelling interaction on human to human dialogue might accomplish this in two ways. First, familiarity with the type of interaction could reduce many of the technical barriers to effective communication inherent in human-machine interaction, especially for naive or intermittent users; in particular, entirely new modes of communication may not need to be learned. Secondly, conversation is actually effective at achieving goals of the sort desired in information seeking via an intermediary; this makes one think that it might therefore be a good model for achieving goals of this sort when the intermediary is a machine. Below, I discuss several approaches to the modelling and study of conversation. It is my hope that this rather cursory review will indicate some of the ways in which conversation might usefully serve as a model for human-machine interaction for information retrieval.

Models of conversation

Conversation, discourse, dialogue, talk are phenomena which have attracted interest in a number of fields, for a variety of reasons. This interest has led to some tentative models and theories of the phenomena, and to some methods for investigating them (for a general discussion, see Coulthard, 1977). I would like briefly to characterize some of these theories and methods. I choose examples from philosophy, sociolinguistics and cognitive psychology, and some related discussion from artificial intelligence. I would like to stress that these are only examples, and that they do not by any means exhaust the variety of approaches to the topic, but represent, at least to me, some extremes of approach.

Three philosophers have greatly influenced the contemporary study of conversation and discourse: J.L. Austin (1962), with 'performatives'; John Searle (1969), with the theory of 'speach acts', and Paul Grice (1975), with his lectures on 'conversational implicature'. For brevity's sake, and because of its emphasis on conversation, I will comment only on the latter's work, although Austin's and Searle's ideas are fundamental to most discourse analysis.

Grice, in a bizarrely disseminated document which became extremely influential without ever having been conventionally published, suggested that there is a logic to discourse (that is, conversation), which follows from some natural rules for conversation. The problem he was concerned with was the alleged discrepancy between formal logic and the 'natural logic' of natural language. In particular, he meant to demonstrate that this discrepancy was only apparent, and arose from a misunderstanding of the 'nature and importance of the conditions governing conversation'. That is, he was concerned with providing a logical basis for explaining how people understand conversations in which sequences of statements do not seem to follow 'logically'. To this end, Grice first formulated a principle basic to a particular type of conversation, the Cooperative Principle (CP):

> "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."
> (Grice, 1975, p. 45)

In accordance with the CP, Grice proposes four categories: Quantity, Quality, Relation and Manner, under each of which fall certain maxims and submaxims. His claim is that these maxims, together with the CP, are necessary, and understood to be necessary, in cooperative conversations. Briefly, the categories and their maxims are:

QUANTITY:
a. Make your contribution as informative as is required for the current purposes of the exchange;
b. Do not make your contribution more informative than is required.
QUALITY: Supermaxim - Try to make your contribution one that is true.
a. Do not say what you believe to be false;
b. Do not say that for which you lack adequate evidence.
RELATION:
Be relevant
MANNER: Supermaxim - Be perspicuous.
a. Avoid obscurity of expression;
b. Avoid ambiguity
c. Be brief (avoid prolixity);
d. Be orderly.

A participant in a conversation, according to Grice, may fail to fulfill a maxim in various ways, including:
a. S/he may quietly and inconspicuously violate a maxim (seeking to misinform);
b. S/he may opt out from a maxim and the CP;
c. S/he may be faced by a clash between two maxims;
d. S/he may flout a maxim.
It is the last case which concerns Grice, for it is this instance that gives rise to 'conversational implicature'; that is, that exploits a maxim for some communicative purpose.

For example, a flouting of the first maxim of quantity:

> "A is writing a testimonial about a pupil who is a candidate for a philosophy job, and his letter reads as follows: 'Dear Sir, Mr. X's command of English is excellent, and his attendance at tutorials has been regular. Yours, etc." (Gloss: A cannot be opting out, since if he wished to be uncooperative, why write at all? He cannot be unable, through ignorance, to say more, since the man is his pupil; moreover, he knows that more information than this is wanted. He must, therefore, be wishing to impart information that he is reluctant to write down. This supposition is tenable only on the assumption that he thinks Mr. X is no good at philosophy. This, then, is what he is implicating). (Grice, 1975, p. 52)

Grice's lectures are of interest in many ways, but two are particularly of interest in our context. First, he has begun to codify the rules by which certain types of conversation (cooperative) are assumed to take place by participants in such conversations (at least in a particular cultural context). And second, he demonstrates that conscious violations of the rules are regularly used by participants to further communication. It is of special interest to note that many such violations, if not all of them, are context dependent (either linguistic or social context or both).

Grice's work has been influential in a variety of fields, especially linguistics and philosophy, but also psychology and sociology. Rather than follow up this influence, I would like to go on to another type of analysis of conversation, that which is now associated with the 'ethnomethodological' schools of sociology and sociolinguistics. Much of this work has strong associations with proposals by Goffman, but it is also heavily dependent upon proposals in the 1960's for an 'ethnography of speaking'. Labov, Sacks and Schegloff are important investigators in this field, and some of their work will be mentioned later, but I would like to characterize it on the basis of a polemical introduction to the general field by Hymes (1972).

Hymes (1972) makes the case for the importance of descriptive analyses of language in a variety of communities. In particular, he, and others of this school, are concerned with the attitudes and knowledge of members of the community, as manifest in the rules by which speakers associate particular modes of speaking, topics or message forms with particular settings and activities. Thus, they are concerned with communication behaviour in its social setting, particularly from the point of view of language behaviour. Again, this means an emphasis on human groups, rather than on grammar per se, but language is still basic data. In particular, these people are concerned with language function; that is, with separating what is said from what is done, and investigating the latter and connecting it regularly with the former. This, of course, is strongly related to Austin's performatives and Searle's speech acts. The emphasis in this school is also highly empirical, indeed Hymes (1972) and Sacks (1972) each suggests that descriptive theory or descriptive analyses and taxonomies are a primary aim of this type of sociolinguistics. Hymes, in this paper, suggests a categorization of units for the study of linguistic behaviour in this context. It is an hierarchical analysis which specifies context and variables, and which states that a complete description of language function must investigate them all.

The highest order unit, the social unit of analysis, is called the speech community. He defines this as a community, sharing rules for the conduct and interpretation of speech and rules for the interpretation of at least one linguistic variety. The next order unit is the speech situation; that is, situations associated with (or marked by the absence of) speech. Examples are ceremonies, interviews, fishing trips, and so on. A speech event is an activity within a speech situation governed by particular norms for the use of speech. For instance, a party is a speech situation, and a conversation during the party a speech event. A speech act is the minimal term within an event, as, for instance, a joke within a conversation. He suggests that discourse may be viewed in terms of acts both syntagmatically and paradigmatically; that is, as sequences of acts, and as choices among classes of acts. Here, a potential conflict with speech act theory arises. Speech styles and ways of speaking are also suggested as categories for investigation, but are not as well defined. More specific is Hymes' tentative categorization of the components of speech.

These components number 16, which I will list: message form (the structure of the message); message content (e.g. topic); setting (physical circumstances); scene (psychological circumstances); speaker; addressor; hearer, or receiver, or audience; addressee; (these four grouped as participants); purposes - outcomes; purposes - goals (these two being differentiated according to community and participant purposes); key (manner or tone of event); channels (medium of transmission); forms of speech (e.g. dialects); norms of interaction; norms of interpretation; and, genres. These, he claims, are the variables of interest in the study of speech in its social context.

Notice that Hymes does not, at least explicitly, suggest a theory of speaking, but rather a method for its investigation. Of course, there is a strong theoretical component in the specification of method and choice of problem, but the eventual statement of generalities follows from the data which are collected. These data and their analyses can be very interesting and fruitful indeed. Hymes' suggestions have been concerned with the study of language behaviour in general. His requirements are so difficult to meet, that there has been no full-scale attempt to cover a speech community in this way. Some people, however, have tried to use his schema, or something like it, in the study of conversation in particular. There have been two especially interesting approaches to this problem, one represented by work by Labov, the other by a school of sociologists taking their lead from Sacks, and now represented by Schegloff in the USA and Coulthard in the UK, among others.

Labov's (1972) orientation has been to try to discover the rules governing the structure of conversation by deep analysis on all dimensions of some conversation. In order to do this, he has proposed some basic units of discourse, and has tried to discover how they are put together in conversation (of a particular sort) and how people manage to understand them. In his search for rules, Labov especially looked at 'questions' and 'answers', of all types. For example, in the sequence:

A: Are you going to work tomorrow?
B: I'm on jury duty.

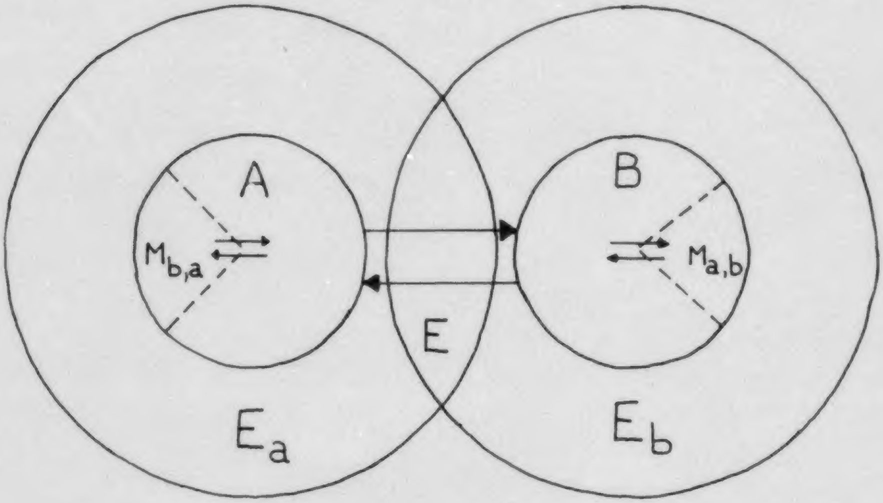
his rule is: if A makes a request of B of the form Q(S1) and B responds with a statement S2 and there exist no rules of ellipsis which would expand S2 to include S1, then B is heard as asserting that there exists a proposition known to both A and B of the form - if S2, then (E)S1, and from this proposition there is inferred an answer to the request (where (E) is an existential operator). This rule is invariant, and emphasizes the role of shared knowledge in conversation. Labov's units and rules are of great interest, especially because of his concern with questions and answers. I do not wish to summarize his work here, but would like to mention its embodiment in the book Therapeutic Discourse (Labov and Fanshel, 1977), in which, by an extremely detailed analysis over a ten year period of one 15 minute conversation between a psychiatric patient and her therapist, Labov and Fanshel manage to formulate a set of general rules for conversation.

Schegloff's work (e.g. Sacks, Schegloff and Jefferson, 1974) is different in that he and his colleagues have been concerned with analysis of one aspect of conversation at a time; typically, with the organization of turn-taking and other sequential structures. The structure that Labov investigated was logical (in some sense); the structure that this school investigates is organizational. The methods also differ, as Schegloff looks at many conversations, rather than one, but the results are equally interesting. Jefferson (1972), for instance, notes a feature which she calls the side sequence, which is a special sort of interruption of a conversational turn, marked by a questioning repeat. That is, something is said by A, and B repeats it. This mechanism is part of a misapprehension sequence, which follows a regular pattern, which is certainly familiar to those of us who have worked on reference desks or as intermediaries.

These studies are examples of a specifically interactional approach to language, concerned with functions rather than, or only in conjunction with words, rather than with the words themselves.

Cognitive psychology has offered a rather different approach to the study of discourse and conversation than those we have looked at so far. Much of the work in this field has been concerned with reading and understanding texts as discourse units, which is rather different from conversation, although certainly related to it. The relation holds especially in the role of shared knowledge, which has surfaced in both the philosophical and sociological approaches to conversation, although in rather different ways. Some cognitive psychologists have concerned themselves explicitly with conversation as communication, and I would like to discuss this approach on the basis of work by Erik Hollnagel (e.g. Hollnagel, 1978, 1979), who contributes to information science as well as to psychology.

Hollnagel's cognitive view is basically that the individual's knowledge (that is, her/his model of the world) is an active mediator in the individual's interactions with the world. In the communication situation, this can be shown in the following diagram of communication between two parties.



$$A \supset M_{b,a} \supset M^2_{a,b} \qquad B \supset M_{a,b} \supset M^2_{b,a}$$

Figure 3. A cognitive communication model (Hollnagel, 1978, p. 7)

What Hollnagel is concerned to demonstrate in this schema is the significance of the environment of each participant, of their shared environment, and of their explictly non-shared environment (environment taken to include, for instance, knowledge). In this schema, one explicitly recognizes that for communication to take place, each of the parties must have a model of the other, which almost certainly does not correspond to the other's model of her/himself. In Hollnagel's analysis, one of the functions of communication can be to try to influence the model that the other participant has of one, explicitly for the purpose of making the communication more effective (at least from the sender's point of view). Hollnagel summarizes the conditions necessary for human communication as:

a. The participants must have a common language or a common code.

b. The participants must have a common understanding of essential parts of the environment.

c. Each participant must have a model of the partner.

Hollnagel goes on to consider the third requirement in some detail, especially as it concerns the notions of credibility and intention. Credibility can be affective or cognitive; that is, based on trust or on knowledge, respectively. Although both are important in human to human communication, Hollnagel suggests that the former is not relevant to human-machine communication, and so concentrates upon describing the essentials of establishing cognitive credibility. One of the most important means to this end is that the partner is able to explain the structure of what s/he is attempting to communicate, and to explain why it is so. This explanation proves to the other partner that the sender does understand the subject matter (I hope that I'm managing this right now, but have my doubts). Another means to the end is by demonstrating, again usually by explanation, competence in the knowledge that is assumed to be shared by the participants. This type of communication Hollnagel calls secondary communication. That is, it is a sequence of questions and answers which are designed to increase credibility, and thus to corroborate the primary communication, the impartation of the information. Notice that secondary communication is concerned with affecting the model that one participant has of the other. Thus, in this sense, such models appear to be necessary for effective primary communication. In Hollnagel's model, the goal of primary communication is to extend the range of the shared environment, which cannot, however, be extended to include the models that the two parties have of one another.

Hollnagel also suggest that knowledge of purpose and intention by each party of the other is useful in dialogues which are aimed at a single goal (that is, where one party is primary). Thus, another function of secondary communication is to establish, in the model that A has of B, B's purposes and intentions in having engaged in the dialogue. The lack of such a model will seriously inhibit the successful conclusion of the dialogue in attainment of the goal. Hollnagel has suggested that one way to model purpose or intention is to understand the user's strategy in problem solving. Since Hollnagel works at the Riso National Laboratory in Denmark, the example of communication between human and machine at Three Mile Island is not inappropriate here. In that situation, one of the primary causes of the disaster was the unwillingness of the human operators to believe the information that the machines provided for them. This information was literally incredible: the suggestion is that perhaps it was of this status just because the machine lacked credibility, and that the cause of this was the lack of a model of the user.

Hollnagel's point of departure in his investigations is to understand, and subsequently model, effective primary communication among human beings (that is, communication of information for specific goals), in order to design effective human-machine communication systems. His emphasis, notice, is always on knowledge, and models of knowledge and participants, and especially on how these models mediate the process of information assimilation. This, then, is an abstract, cognitive view of communication, with some, but not much, bearing and interest in language itself.

Computational linguistics and artificial intelligence are fields which are specifically concerned with language; the former strictly so, the latter primarly in its attempt to simulate intelligent behaviour. I will not concern myself with the question of natural language understanding systems, which has been a primary concern in these fields, but rather with investigations in other aspects of the human-machine interface in human-machine systems. A number of people have concerned themselves with this problem, especially with the issues of: to what extent these interfaces ought to reflect human-human conversational modes; and how the human-machine dialogue is to be implemented (see, e.g. ASSOCIATION OF COMPUTATIONAL LINGUISTICS, 1980). Hayes (Hayes &

Reddy, 1979; Hayes, 1980) is representative of the school that says that, in order for human-machine communication to be effective, it will be necessary for the systems to incorporate a number of aspects of human dialogue structure. He calls these the 'non-literal aspects of communication', arising from

> non-grammatical utterance recognition;
>
> contextually determined interpretation;
>
> robust communication procedures;
>
> channel sharing. (Hayes, 1980)

The first two points are relatively clear, but the last two require some clarification. By robust communication procedures, he means such things as error correction, back-tracking, asking for clarification, etc. By channel sharing he means turn-taking, in the sense of Sacks, Schegloff, and Jefferson (1974). Hayes' contention is that these aspects of dialogue have not been taken into account in most human-machine interfaces, and when they have been considered the results have been

> a. naive, e.g. half-duplex mode, brooking no interruptions;
>
> or,
>
> b. straightforwardly imitative of superficial aspects of non-literal communication, e.g. 'Hi there, I'm your friendly computer system'.

In particular, Hayes suggests that many aspects of non-literal communication can now be incorporated in human-machine systems in ways that are not possible in human-human communication, but which are appropriate to the human-machine environment, and are responsive to dialogue functions. This is largely due to technology of interaction, and he bases most of his argument on hardware which allows simultaneous presentation of a variety of information, with provision for highly robust means of interaction. Thus, systems can be designed which indicate, on the screen, what the machine considers to be the current focus of attention, or where the machine is having problems in understanding. Similarly, many options can be presented simultaneously for choice by the user (as in a menu system), if clarification is necessary (cf. *side sequences*). He suggests that, because of technology which allows multiple output to the screen while the machine is still accepting (and echoing) input, turn-taking in the human sense may not be necessary at all. And so on.

The point I want to make here is that Hayes (and others, e.g. Schneiderman, 1980; Hiltz, 1980) are suggesting that the modes of communication possible between human and machine are such that human dialogue structure should *not* be mimicked, since it is no longer directly appropriate, but rather that new structures can be devised which respond to the communication *needs* and *functions* represented by human dialogue structure. Unfortunately, and this is opinion, it seems to me, and others (e.g. Schegloff, 1980; Wynn, 1980), especially sociologists, that artificial intelligence's understanding of these needs and functions is perhaps not too well informed, and often biased by the attractiveness of the technology.

The relevance of conversation models to information retrieval

I have presented examples of four different approaches to the analysis and modelling of conversation: philosophical/logical; interactional; cognitive; and, technical. Because each is grounded in a particular intellectual framework and in a particular problem structure, each is necessarily context dependent, and therefore not applicable directly and *in toto* to some other situation (e.g. the information retrieval context). I would like to point out, briefly, some ways in which each of these fails in its application to information retrieval.

Recall again the information retrieval context. It is a formal, goal-oriented communication system, with two parties, one of whom, the user, instigates the communication and evaluates its success, the other, the intermediary, functions as expert in the organization and manipulation of the data base. Each participant has specific tasks to perform if the entire communication is to be as successful as possible, but there are certain barriers to their accomplishment. In particular, it is difficult for the user to indicate to the intermediary what is necessary to resolve the problem, and socio-cultural factors mitigate against straightforward achievement of goals.

The philosophical/logical approach to conversation has little to say about the interactive aspects of this situation, nor does it deal well with the question of relative roles. The interactional approach has been unable to say much about the cognitive, that is, topic and content related aspects of conversation. The cognitive approach pays relatively little attention to the linguistic data. And the technical approach has little understanding of the dialogue itself.

On the other hand, all of these approaches to dialogue have something in common, and that is the emphasis which they all place on the importance of shared knowledge to effective communication, and in their insistence that it is the functions rather than the surface of the dialogue that are significant. Although it is not much, and somehow not startlingly new, this kernel of agreement could, I think, serve as a basis for the design of effective, conversationally based interfaces for human-machine information retrieval systems. How could we go about achieving this aim?

We have, fortunately, available to us in this quest, some excellent data for analysis. My suggestion, and indeed it has already been taken up in various ways by several investigators (e.g. Ingwersen & Kaae, 1979; Mark Pejtersen, this volume; Lynch, 1978), is that we use the data of the reference interview, or the user-human intermediary interaction in order to:

a. understand the particular rules and behaviour of information retrieval dialogue and its functions; and,

b. develop a model which could simulate the effective aspects of this dialogue in a user-machine intermediary system.

It seems to me that this could be accomplished by combining the methods of analysis and the normative structures of the approaches to dialogue that I have mentioned (to include those that I have not mentioned as well: e.g. Gordon & Lakoff, 1971; Brazil, 1975), informed by what we already know of the situation. Thus, we know that achieving the desired shared knowledge is difficult for some specific reasons - we can use this knowledge to inform our analysis of the dialogue by, for instance, looking for aberrations in the normal, 'logical' pattern of information retrieval conversation. We know, indeed, that the situation is not the same as normal, or ordinary unmarked conversation. The results of the interactional approach tell us that the speech situation is important, and must be described in specifically social terms, and suggest how this might be done. The credibility, or lack of it, of the human intermediary, and indeed of the user, investigated as to its causes, would inform us of what functions the machine intermediary must perform. And the implementation of the functions of the human intermediary in a machine intermediary might be achieved through the insights of the technical approach.

Conclusion

So, my suggested program of research is:

1. User-machine interaction in the information retrieval situation ought to be dialogue structured, because of the nature of the information retrieval situation;

2. The example of the reference interview or user-human intermediary interaction provides excellent data for analysis leading to a model for this dialogue;

3. The analytic and normative techniques of the various forms of discourse analysis, while not sufficient to the problem on their own, together and selectively could be used as the basis for such analysis;

4. Therefore, before designing more user-machine interfaces for information retrieval systems, data should be gathered from the natural setting of information retrieval dialogue and analyzed to discover the functional rules and structures of this type of conversation.

The mix of techniques and order that I see for this program is: first, to gather the data (and here I suggest, depending on specific goals for the individual investigation, either deep analysis of all of one or a few interactions, or detailed analysis of one aspect of many interactions; second, analyze the material initially in terms of the Cooperative Principle; third, consider those instances where the Cooperative Principle is apparently violated and investigate them interactionally; fourth, identify and characterize the rules for credibility assessment and knowledge sharing; fifth, develop a normative set of rules; and, finally, use these to reanalyze the situation and to develop the rules.

It is not much of a framework, yet, but I think it has promise. I know several projects already working within this general framework, so it seems to me not entirely without merit. My hope is that the cooperative efforts of various researchers working on various aspects of the information retrieval dialogue will result in a general model of information retrieval dialogue which will be usable for (among other purposes) the design of a new type of truly effective interactive interfaces for human-machine communication for information retrieval.

References

1. ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (1980). Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics. Philadelphia, ACL.

2. AUSTIN, J.L. (1962). How to Do Things with Words. Oxford, Clarendon Press.

3. BELKIN, N.J. (1980a). The problem of matching in information retrieval. In: Theory and Applications of Information Research. London, Mansell: 187-197.

4. BELKIN, N.J. (1980b). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, v. 5: 133-143.

5. BELKIN, N.J., BROOKS, H.M. & ODDY, R.N. (1979). Representation and classification of anomalous states of knowledge and information for use in interactive information retrieval. In: IRFIS 3. Oslo, The Norwegian Library School: 146-183.

6. BELKIN, N.J. & ROBERTSON, S.E. (1976). Information science and the phenomenon of information. Journal of the American Society for Information Science, v. 27: 197-204.

7. BRAZIL, D.C. (1975). Discourse Intonation. Discourse Analysis Monographs 1. Birmingham, English Lanuage Research.

8. COULTHARD, M. (1977). An Introduction to Discourse Analysis. London, Longman (reprinted 1981).

9. GOFFMAN, E. (1955). On face-work: an analysis of ritual elements in social interaction. Psychiatry, v. 18: 213-231.

10. GORDON, D. & LAKOFF, G (1971). Conversational postulates. In: Papers from the 7th Regional Meeting of the Chicago Linguistic Society. Chicago, CLS: 63-84. Reprinted in: Syntax and Semantics 3: Speech Acts, P. Cole and J.L. Morgan (eds). New York, Academic: 83-106.

11. GRICE, H.P. (1975). Logic and conversation. In: Syntax and Semantics 3: Speech Acts, P. Cole and J.L. Morgan (eds). New York, Academic: 41-51.

12. HAYES, P. (1980). Expanding the horizons of natural language interfaces. In: ACL (1980). Parasession on Topics in Interactive Discourse: 71-74.

13. HAYES, P. & REDDY, R. (1979). An anatomy of graceful interaction in man-machine communication. Technical Report, Computer Science Department, Carnegie-Mellon University.

14. HILTZ, S.R., JOHNSON, K. & RABKE, A.M. (1980). The process of communication in face-to-face vs. computerized conferences: a controlled experiment using Bales interaction process analysis. In: ACL (1980). Parasession on Topics in Interactive Discourse: 75-79.

15. HOLLNAGEL, E. (1978). Qualitative aspects of man-machine communication. Report Riso-M-2114. Roskilde, Denmark, Riso National Laboratory.

16. HOLLNAGEL, E. (1979). The relation between intention, meaning and action. In: Informatics 5: The Analysis of Meaning, M. MacCafferty and K. Gray (eds). London, Aslib: 135-147.

17. HYMES, D. (1972). Models of the interaction of language and social life. In: Direction in Sociolinguistics, J.J. Gumperz and D. Hymes (eds). New York, Holt, Rinehart and Winston: 35-71.

18. INGWERSEN, P. & KAAE, S. (1979). User-librarian negotiations and information search procedures in public libraries: an analysis of verbal protocols. In: IRFIS 3. Oslo, Norwegian Library School: 71-106.

19. JEFFERSON, G. (1972). Side sequences. In: Studies in Social Interaction, D. Sudnow (ed). New York, The Free Press: 294-338.

20. LABOV, W. (1972). Rules for ritual insults. In: Studies in Social Interaction, D. Sudnow (ed). New York, The Free Press: 120-169.

21. LABOV, W. & FANSHELL, D. (1977). *Therapeutic Discourse*. New York, Academic.

22. MARK PEJTERSEN, A. (this volume). Theory and practice of the librarian's role as a mediator in fictive literature.

23. LYNCH, M.J. (1978). Reference interviews in public libraries. *Library Quarterly*, v. 48: 119-142.

24. ODDY, R.N. (1975). *Reference Retrieval Based on User-Induced Dynamic Clustering*. Ph.D. Thesis, University of Newcastle-upon-Tyne.

25. SACKS, H. (1972). An initial investigation of the usability of conversational data for doing sociology. In: *Studies in Social Interaction*, D. Sudnow (ed). New York, The Free Press: 31-74.

26. SACKS, H., SCHEGLOFF, E.A. & JEFFERSON, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, v. 50: 696-735.

27. SCHEGLOFF, E.A. (1980). What type of interaction is it to be? In: *ACL* (1980). Parasession on Topics in Interactive Discourse: 81-82.

28. SCHNEIDERMANN, B. (1980). Natural vs. precise concise languages for human operation of computers: research issues and experimental approaches. In: *ACL* (1980). Parasession on Topics in Interactive Discourse: 139-141.

29. SEARLE, J.R. (1969). *Speech Acts*. London, Cambridge University Press.

30. WYNN, E. (1980). What discourse features aren't needed in on-line dialog. In: *ACL* (1980). Parasession on Topics in Interactive Discourse: 87-90.

LANGUAGE AS A BARRIER TO COMMUNICATION

Ralph Adam

> "When _I_ use a word," Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean - neither more nor less."
> "The question is," said Alice, "whether you _can_ make words mean different things."
> (Lewis Carroll, _Through the Looking Glass_).

In my 1979 paper[1] I outlined some of the communication problems which beset social science research and suggested ways in which improvements might be made. My theme was that information services should reflect the information needs and habits of their likely users: there is little point in supplying services if we have no ideas as to how, if at all, they are used. It is also unfortunate that much information research is disseminator-orientated.

One area which I considered was that of language. In this paper I should like to develop these ideas a bit further.

The role of language is generally taken for granted in the information world. This is partly a reflection of the 'communication research' ideology. It must also be due to the paucity of knowledge about the part language plays in scientific creativity: very little is known about its effect on the growth of knowledge.

It has been fashionable for at least twenty years to criticise the way in which social scientists use language. We are all familiar with the sort of casual comments (sometimes in 'formal' places) that are made: information specialists complain that they cannot supply effective services because of social scientists' insistence on using woolly language; non-specialists (including other scientists) complain about writers who cloak common-sense observations in impenetrable jargon; there have even been several strong attacks from within the social science community.

Some of these criticisms are justified, while others are made for political or other personal reasons, such as a desire to score points. Many, however, are based on ignorance or a failure to understand the material by people who do not speak the appropriate professional language.

It is a commonly-held belief that, because the social sciences are concerned with the study of everyday behaviour, their methods and findings must be capable of being understood by any intelligent layman. This is one cause of the poor quality of many studies of information use: they are carried out by people with no training in social research who feel that question-asking is a simple process and that the results can be made 'scientific' by applying a few statistical techniques. This is an extremely naive view of what is involved in sociological or psychological research and reflects an attitude that was lambasted by Sorokin[2] as long ago as the fifties.

Similarly with language: we are all experts on human behaviour, so there is no need to disguise it with mystifying language. Such a view often tells us more about the narrow-mindedness of the critic than it does about the subject of his criticism. Indeed, the casual use of terms from sociology, psychology or economics by the uninitiated can give rise to the very same linguistic confusions which are complained of.

I do not want to be seen as a defender of bad language. What I should like to suggest is that the use of language is a complicated issue about which we know very little. Language plays an important part in social science research and yet its role is often underplayed. It can be as much a barrier to communication as an aid to it. Until we know more about linguistic practices and their background we are in a poor position to understand the findings of research on users and we cannot be considered competent to provide them with effective services.

I have dwelt on our failure to appreciate how language is used because it seems, in part, to be a reflection of research methods in information science. Two main approaches have been used to study communication behaviour. One can be called the structuralist view. This tradition relies on an objectivist stance and treats communication as a series of discrete and quantifiable operations. Those carrying out such research have not always been social scientists and many have lacked an awareness of the variability of human behaviour. Into this category come most of the studies which have looked at the structures of communication: investigations of library usage or journal readership, studies of selected authors' or conference participants' backgrounds and so on.

Some of this work provides us with useful data, but much of it is of doubtful value and falls into what Gellner has called the Cargo Cult approach to research.[3] He suggests that the apparent success of natural scientists has been so great that they can hardly be considered to be of the same species as the rest of us. In much the same way as when technologically-primitive cultures are suddenly brought into contact with the overwhelming wealth and superiority of another civilisation and make hysterical attempts to create short cuts to attain this alien wealth, so some social researchers try desperately to imitate the methods of natural scientists in the hope of magically transforming their work into 'respectable' science. (It is not only in the user studies area that this applies: a recent U.S. study of male sexuality made sweeping generalisations on the basis of a few statistics and a questionnaire of which 119,000 copies were distributed and only 7,000 returned.[4] This type of sex research is treated with amusement, yet when information research is done so badly the results are taken seriously).

The emphasis placed on the structuralist approach to information research has been one reason for the lack of consideration given to the role of language. This type of research makes use of some sociological techniques, but the resultant work is not sociological.

The second research style takes an interactionist approach and owes much more to sociology. Into this category can be put those studies which have investigated issues from the user's point of view. Relatively little information research has been carried out within this tradition and it is sometimes dismissed as 'unscientific' because of its qualitative, rather than quantitative, emphasis. This feature can also reduce the chances of obtaining funds for research.

These two approaches to research illustrate the distinction between 'user' and 'use' studies. The ideal method probably owes something to both and the dual approach can be seen in the type of research which has recently been carried out by Tom Wilson and Bill Blake.

I should like now to consider some of the issues which relate to social scientists' use of language. It is important to remember the distinction between those linguistic features that are essential to, and an integral part of, the work of a discipline, and those practices which are not entirely necessary and can lead to confusion or inhibit communication in some other way. If most of my examples come from sociology, this is partly because it is my own discipline (and I am more familiar with its problems) and partly because it is the language of sociology which receives the most criticism.

An important feature of the social sciences is the role of theory. Sometimes this is indistinguishable from certain forms of philosophy - much work is carried out at a conceptual level and there is an emphasis on the analysis of words and their meanings. The importance of theoretical work is often ignored (or, at least, under-played) by non-specialists as well as by those working in more 'applied' fields. Nevertheless, good research in the social sciences cannot exist independently of theory. The same would probably apply in any other scientific discipline.

There is however a conflict between the need for (good) theoretical work and its inaccessibility to non-specialists, and this can lead to problems. The findings of social science research need to be directed at a very wide range of readers: sponsors (including governments), journalists (and, through them, the general public), professionals in areas such as social work, education and the law. Where a topic such as race relations or crime is concerned there may be a need to produce quick results. Yet, while most of the audience will probably be familiar with the subject matter of the work, it is unlikely that many will have the background to appreciate the theory. This leaves authors with a choice: they must either present their writing in specialist language and risk not getting the message across, or else water the material down to an everyday level and take the other risk that important aspects will be lost or that ideas may be taken out of context.

In highly topical research the need for fast results means that there is great pressure on researchers to produce material which can be consumed rapidly by interested parties. Social science research, by its nature, has a high verbal content. It follows that, if there is a need to produce something quickly, there may not be the time to create the best quality written product. This contrasts with the natural sciences where the stereotype is of a chemist or physicist doing mysterious things in his laboratory. It is accepted that not all his experiments will end in earth-shattering results. Indeed, they may not be successful at all. And it is not taken for granted that we will all be able to understand such findings as do result. Social science research, on the other hand, is assumed to be open to the comprehension of all. There have been a few writers who have been able to present their work to general audiences (R.D. Laing, J.K. Galbraith and H.J. Eysenck come to mind): they have mostly been exceptional people and their very ability to popularise carries with it the danger that they may obscure the work that has gone into carrying out the original research. There is a great difference in any discipline, between writing for a scientific audience and for a general one: the average reader of a popular work can have little idea of the amount of research that lies behind it.

A further danger of popularisation is that ideas may be removed from their contexts. Social science languages contain many words taken from everyday usage, but with specialised meanings. Such terms can easily be picked up by journalists and others who fail

ERIC Clearinghouse on Information Resources
School of Education
Syracuse University
Syracuse, New York 13210
(315) 423-3640

to appreciate their technical meanings and who may debase the terms by using them in mistaken contexts. This risk can, to some extent, be avoided if new terms are created to describe such everyday situations. This, however leaves researchers open to criticism for inventing neologisms to describe phenomena which are already well-known. Specialist terms lose their value if they become incorporated into ordinary language and I shall say more about this later.

I have been referring to 'social science languages', but there is really no such thing as the language of sociology or the language of economics. Because the social sciences tend to be perspective-,as well as discipline-based, each theoretical perspective has its own terminology. Thus, while the writings of a marxist or a symbolic interactionist are recognisable from the words they use, the work of one group may not be comprehensible to members of others and, for example, different terms may be used to describe what are apparently the same phenomena. This division by perspective is another reason for the seeming proliferation of terms.

The language of the natural sciences has developed an image which suggests that it is value-free and has an almost clinical precision. Some social scientists, particularly those in the positivist tradition, have fallen for this image and feel that their language, too, should be cleaned of its evaluative and emotive connotations. But social behaviour cannot be presented in this way and those who try to give a contrary impression are making the same mistake as are the other followers of the scientific cargo cult.

Julius Gould has distinguished between three types of vocabulary which are to be found in sociology: the technical, the descriptive and the theoretical.[5] These categories, which are not mutually exclusive, can be applied usefully to all the social sciences.

Technical vocabularies are the easiest to discuss. Every science depends on certain operations or techniques which are essential components of its practitioners' tool kits and which permit ease of communication. The social sciences have their own special methods, many of which are well-known. Examples are survey research, personality testing and model building. These techniques have their own vocabularies and an understanding of them is becoming increasingly important for the readers as well as for the authors of research reports. The use of words borrowed from mathematics can give the impression that such work has a 'hardness' and objectivity that other social science research lacks and this tends to create a gulf between the supporters of 'technical' research and those who subscribe to more theoretical approaches. Such research appears less value-free, however, when the statistics come to be interpreted!

The issue of descriptive language is rather more problematical. By 'descriptive' I mean that work which is concerned with describing social, economic or political institutions. These are the areas that affect daily life and for anyone studying them there is a ready-made vocabulary available. This is where the confusion can arise: a ready-made vocabulary can imply ready-made solutions. Because we are all familiar with terms like 'family', 'capitalism' or even 'library', we feel that we ought to be competent to discuss the problems that surround them: even if these problems are complex. The term 'family', for instance, can relate to kinship patterns, household units or a set of relationships. Some terms, of which 'democracy' is a good example, can have a multitude of definitions - in this case, almost as many as there are writers on the subject. Other terms ('tolerance'- of drugs, for instance)[6] have technical meanings which can be quite distinct from their everyday ones. I have already referred to the problems of terms with multiple meanings. The converse situation, where several terms relate to the same concept, can also lead to confusion.

New terms may be invented not only by authors who wish to preserve their work from contamination by meanings imputed from daily life but also for more aesthetic reasons. For example, the term 'bourgeoisification' was introduced by a team studying English social mobility to refer to the adoption of middle-class values by well-off members of the working class. They subsequently decided it was a rather ugly term and changed it in their later publications to 'embourgeoisement', with its more cultured French air. It was the latter term which came into general use, even though its English connotation is rather different from the original French one. Different terms can also apply to the same, or similar, concepts if they have been developed by members of different theory groups who do not read or understand one another's work.

The third type of vocabulary is the theoretical. This is by far the most difficult to control. Theorists need accurate language just as much as do technicians. Yet, the theoretical vocabularies of the social sciences are full of confusions. One reason for this is that there is no theory which is sufficiently comprehensive or over-arching to encompass all topics. Consequently, 'fuzzy' areas develop at the boundaries where, unlike the physical sciences, little consensus exists. Also, very few theories are able to transcend cultural barriers, irrespective of how universally applicable they may appear to be. Some theories are not theories at all: merely sets of propositions which sometimes fail to add up. So, there are not only problems in search of theories but we also have theories in search of problems. Any behaviour, however irrationel, can be justified if it is couched in sufficiently high-sounding language which few readers can untangle. Unfortunately, there are some theories in the social sciences which seem to boil down to absolutely nothing! And it is these which give the social sciences a name for hiding conventional concepts amongst impenetrable jargon. Gould has even gone so far as to suggest that the word 'theory' is the most frequently used and ill-defined in the social sciences.

The lack of any standard theoretical language for social scientists has led to the development of two schools of thought amongst philosophers of social science. One has seen this as a sign of the under-development (or relative backwardness) of a group of young sciences. This view received support from those members of the Vienna Circle who tried to create a Unified Science in the thirties: a plan which would have required a standard language to permit communication between members of different disciplines. The alternative viewpoint is to see the social sciences as being related to one another, but as having a separate form of development, and a different linguistic structure, from the natural sciences. Many information scientists have favoured the former approach - possibly because its adoption makes their own lives easier.

My comments on theoretical vocabularies may suggest that I am really a supporter of those who see the social sciences as being under-developed. I am not. What I hope I am making clear is that there is a variety of theories, some good and others bad, as well as a range of theoretical perspectives. All of these generate their own terminologies. We must accept that this is a feature of the social sciences. It is a mistake for information specialists to think that they are doing social scientists a service by forcing them to standardise their language. What may appear to the layman to be excessive jargon may, in fact, be a form of expression that has been specifically designed to suit the needs of a particular group. The design of effective information systems should take this into account and the views of users be incorporated before such an enterprise can be successful.

One current approach to easing communication within the social sciences is the movement for conceptual and terminological analysis. I shall say more about this shortly. First, however, it is necessary to establish what lies behind it. The main issues are considered to be two-fold: clarification of concepts and ease of translation.

Attempts to increase the precision of language have generally been the preserve of conservatives. In recent times, such moves have been associated with the Popperian or Kuhnian schools. The arrival of mechanised information retrieval systems has given an undoubted impetus to this activity. However, even strong supporters of the relativist view of scientific development must acknowledge that the current state of information retrieval technology allows for little freedom of linguistic style. A major improvement in this respect has been the creation of Sociology of Education Abstracts which caters for a variety of theoretical perspectives. But it is difficult to see this example developed on a large scale without a means of switching from one mode of language to another. And, for such switching to take place, some form of control is necessary.

One way of doing this is by creating formal links between concepts, words and terms and these need to be distinguished with care.[7]

A concept is a form of mental construct. A little earlier, I mentioned the concept of 'embourgeoisement'. Although the term was first used about twenty years ago, the idea had been discussed for several centuries - certainly during the Scottish Enlightenment. Marx had a few things to say on the topic, as well. However, none of these writers had given a name to the concept. And it is the naming that produces the most trouble. (For the purposes of this discussion, I shall ignore the chicken-and-egg problem of how we can know we are referring to the same concept without having first accepted a common definition).

A term is the name given to a concept in a specified context (otherwise it is merely a word): in the above case 'embourgeoisement'. One concept can have several names and the same name can apply to a number of concepts. We can minimise the problems caused by this through the use of classifications and definitions.

Classification involves bringing ideas together, so that if I say that all lectures in information science are boring, I am classifying information science lectures in a particular way. In biology, it is possible to devise schemes which provide non-intersecting classes into which any individual can be placed. The same cannot be done in the social sciences, where relationships between entities are so complex and variable that they can only be placed in simple (and overlapping) categories. Most biological features are relatively fixed. Thus, if I see a large growth on someone's nose, it is unlikely to be a sign that that person is turning into a rhinoceros; it probably signifies a tumour. On the other hand, if I see a public commotion in a busy street I cannot predict whether it is the start of a full-scale revolution, or merely housewives looting shops because they object to price rises.

Definitions create greater problems, especially in sociology. Definition usually involves the labelling or naming of concepts, so that when a term is used subsequently it has a particular meaning. There is no need for a given term always to have the same meaning so long as we are able to verify it in a particular case. In sociology, however, an added confusion has been caused by those who consider it possible to define things as well as words.

This arose originally through the work of one of the most influential theorists of the inter-war years, W.I. Thomas. He created the idea of the 'definition of the situation' which he used both to describe the way groups set standards through the creation of norms, codes and laws and also to describe the ways in which individuals see their own experiences and sequences of experience. Thus, behaviour is determined by certain conditions which make up the 'situation', including the state of the individual, his environment and the subjective manner in which these are perceived, evaluated and made conscious. The basis of action is subjective, but the result is not. I think it would have been clearer had Thomas spoken of individual descriptions or classifications, rather than definitions, of the situation, especially as much confusion has resulted from this usage. Individual views or definitions of a situation can be mistaken and cannot, anyway, be related to one another.

The problems of translation are familiar in daily life as well as in science. Two approaches are available: both ideas and words can be translated. Translating words is generally the simpler method. It can lead to the type of error that makes for amusing fillers in newspapers during the 'silly season'. It can also produce the turgid translations of scientific papers which seem to take all the interest out of the originals. The translation of ideas requires much more creative effort and is facilitated by the availability of glossaries of technical terms, to ease conceptual clarity. Thus, for example, the meaning of the word 'state' might be distinguished according to context: not just to separate its use in psychology from that in politics, but also in order that its various uses in politics may be distinguished from one another.

In the arts, such translations are particularly difficult. We can imagine the problems a translator of Macbeth might encounter by taking the word-for-word approach to a statement such as: "Out, out, brief candle! Life's but a walking shadow." A technical dictionary would not be much help, even if it could supply translations for words like 'candle' or 'shadow'![7]

A major obstacle to translating the social sciences is that different cultural styles produce different conceptual forms. This can be accentuated when different orthographies are used. Work carried out in one society may mean little in another. Some interesting results have been produced by experiments in 'back translation', in which material is translated by one person from language A into language B and then, by someone else, back into language A.[7] There is a definite need for help with translation as very few English-speaking social scientists have mastered a second language. Translation must be possible, not just between national languages, but also between perspectives, if it is to be provided on the scale necessary.

The conceptual and terminological analysts hope to overcome these problems by creating a conceptual data-bank to generate dictionaries, thesauri, glossaries, an international encyclopaedia of social science concepts and a document classification scheme. The aims of this activity have been given as:

1. to facilitate the conduct of research;
2. to improve scientific and technical writing through improved clarity and precision
3. to improve the understanding of social science results through better understanding of the terms used: within and across scientific communities, by the public and across language and cultural barriers;
4. to improve information storage and retrieval for scientists, scholars, practitioners and the public.[8]

The existence of a movement with such aims is laudable. However, it could easily fall into the trap of providing a focus for the cargo cultists, with their mistaken analogies, while ignoring the most complex issues and limiting its work to the most easily codified theoretical perspectives.

Perhaps an immediate improvement in communication might be achieved by following the recommendations of C. Wright Mills: that every writer in the social sciences should first ensure that he thinks clearly and then present his ideas in as simple a form of language as his subject and his thought about it permit. Indeed, it is something from which we could all benefit. But, attractive as this recipe may sound, it cannot provide an adequate solution to the problems which I have been discussing. It is especially important in a time of economic restraint that services of any sort are effective. This means that they must meet the needs of their users and service suppliers should appreciate what these needs are.

Social science information services are still in their infancy. Before they reach adulthood (and adulthood always seems to come sooner than one expects) it is essential that sufficient research is carried out into users' needs, for mistakes could prove very costly. My intention in this paper has been to demonstrate that linguistic practices should receive a high priority as a research topic because of the important role which language plays in communication.

References

1. R. Adam. Is there a future for social science information systems? In: IRFIS 3: Proceedings of the Third International Research Forum in Information Science, edited by T. Henriksen. Oslo: Statens Bibliotekskole, 1979. pp. 203-220.

2. P.A. Sorokin. Fads and Foibles in Modern Sociology and Related Sciences. Westport, USA: Greenwood Press, 1976. (see esp. chapter 2: "Verbal defects").

3. E. Gellner. The alchemists of sociology. Inquiry, vol. 2, pp. 126-135 (1959).

4. The project referred to is reported in: S. Hite. The Hite Report on Male Sexuality. New York: Knopf, 1980. Although the questions were rather more sophisticated and less likely to cause embarrassment or lead to fantasied replies, the INFROSS project depended on a very low response rate for

its findings. Out of nearly 3.000 questionnaires distributed, only about 1.000 usable replies were received.

5. J. Gould. The vocabulary of sociology. British Journal of Sociology, vol. 14, pp. 29-41 (1963). Reprinted in: The Phenomenon of Sociology by E.A. Tiryakian. New York: Appleton-Century-Crofts, pp. 313-324.

6. I have discussed the use of this word in detail in: R. Adam. Terminology and the social sciences: a case study. Communication and Cognition, vol. 8, pp. 3-12 (1975).

7. F. Riggs. The importance of concepts: some considerations on how they might be designated less ambiguously. American Sociologist, vol. 14, pp. 172-185 (1979). The article is followed by several comments by other writers and a rejoinder by Riggs.

8. The CONTA Conference. International Classification, vol. 8, p. 101 (1981). The work discussed at this conference has been going on for some years under the auspices of the Committee on Conceptual and Terminological Analysis and, latterly, as part of the Unesco Inter-Concept Project. Riggs is one of the initiators of both projects.

ACTION RESEARCH AND USERS' NEEDS

T.D. Wilson and D.R. Streatfield

One of the recurring themes within information science is the application of research results to professional practice. In calling for research proposals, practitioners often seem to demand something which will show usable results quickly, since they perceive their problems as essentially practical. They sometimes find it difficult to grasp that if the research is intended to be fundamental it often cannot be carried out quickly and may not necessarily have practical results.

The difficulty that arises for the researcher is that, on the one hand, the practitioner wants applicable results, and, on the other hand, research funders generally want 'sound research'. At times the researcher is caught in the middle, trying to produce proposals to satisfy both. When research results fail to make any significant impact on the world of practice the response of funding agencies (not only in our field but in others too) is to stress the need for 'more effective dissemination'.

The proposition that dissemination of research results will increase the probability of application sounds reasonable. However, the process appears to be ill-defined and uncertain and the proposition can be doubted where the aims of research are basic rather than applied. It seems to us that there is confusion between the application of research results in science and their application in practice. We will explore this distinction further by reference to figures 1 to 3.

In figure 1 we have tried to illustrate the processes which are assumed to occur in disseminating research results using 'the scientific method model'. The key fact to bear in mind here is

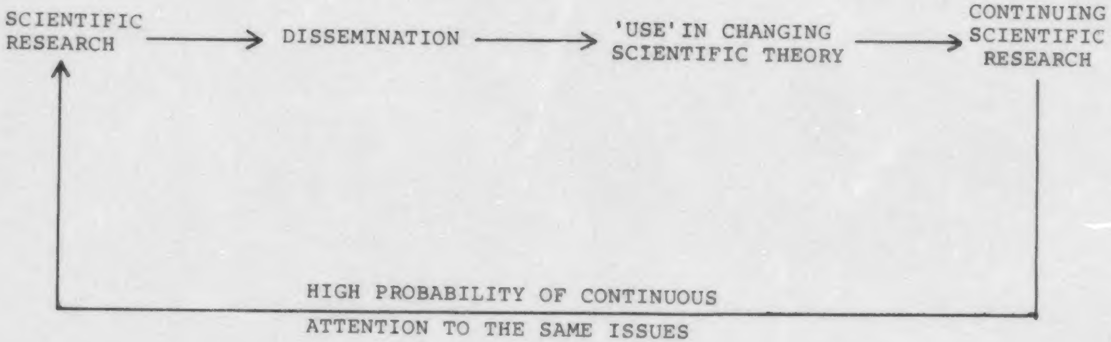

SCIENTIFIC RESEARCH
DISSEMINATION
'USE' IN CHANGING SCIENTIFIC THEORY
CONTINUING SCIENTIFIC RESEARCH
HIGH PROBABILITY OF CONTINUOUS ATTENTION TO THE SAME ISSUES

FIGURE 1

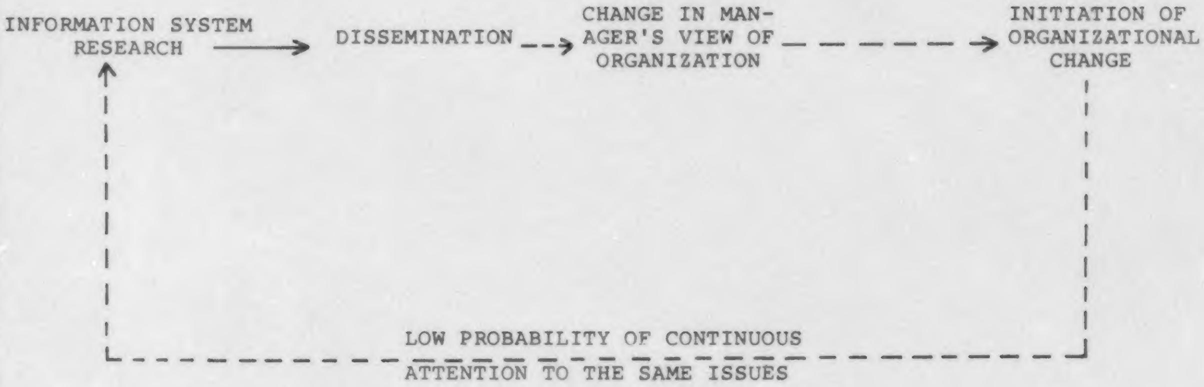

INFORMATION SYSTEM RESEARCH
DISSEMINATION
CHANGE IN MANAGER'S VIEW OF ORGANIZATION
INITIATION OF ORGANIZATIONAL CHANGE
LOW PROBABILITY OF CONTINUOUS ATTENTION TO THE SAME ISSUES

FIGURE 2

that, initially at least, the results of scientific research are applied in science itself rather than in any other field of practice. That is, results are used to support, confirm or deny the propositions of formal scientific theory. The use of science in technology or engineering may come rather later: witness, for example, the discovery of the properties of semiconductors in the 1920s and their application in the development of the transistor in 1948.

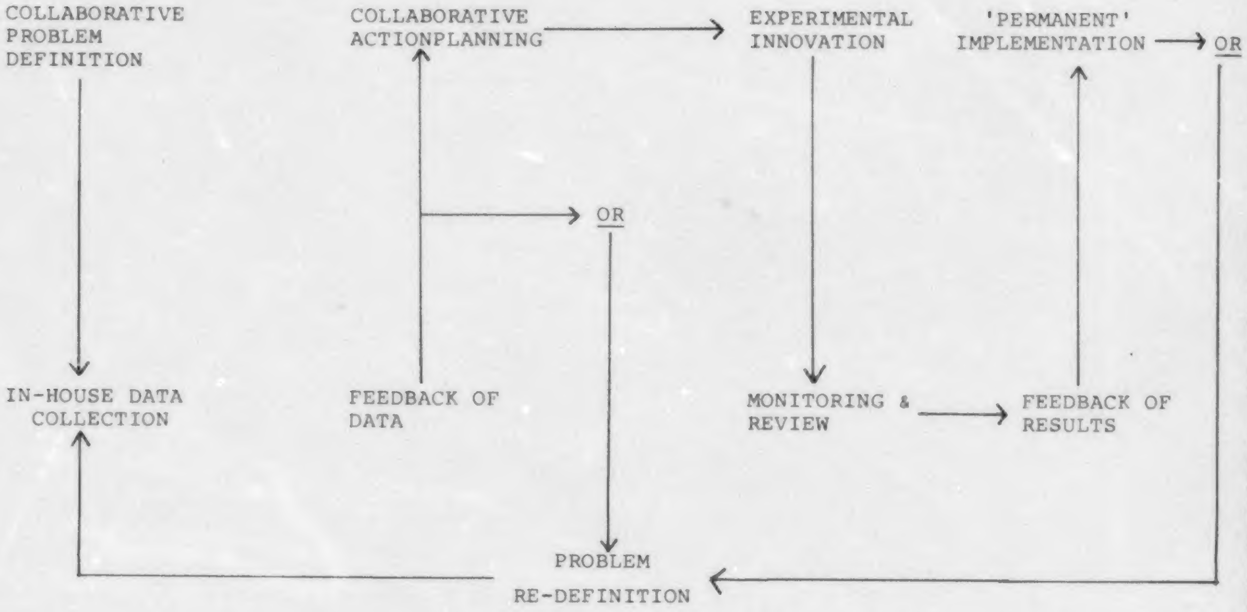
Figure 1 also suggests that, because scientific research finds its application within science, there will be a high probability that virtually continuous attention will be given to the same issues over long periods of time.

In information science, however, the research we are involved in is usually applied research: that is, it is related to problem-solving in the real world. (This is not to deny the possibility of a basic science of information but at present anything basic in our field appears to belong firmly within the bounds of some other discipline and information science takes on more and more the appearance of a multi-discipline). Almost always, too, our research is carried out in a real organizational world in which the application of research results necessarily involves some degree of organizational change.

If we assume that dissemination alone can bring about the necessary degree of organizational change then presumably we have in mind a process something like that outlined in Figure 2. The improbability of this actually happening on any large scale is so glaringly obvious that we have chosen to show it by using a broken line. The reason for the improbability is to be found in the box 'change in manager's view of organization': in other words, a change in attitudes is necessary, and social psychologists are fully aware that attitudes are deeply rooted and difficult to change. The force of a research finding would have to be so overwhelming that not only would the need for change be recognised but the will to change would be stimulated.

If we reject the notion that dissemination of results can work in this miracle way what is the alternative? One possibility

FIGURE 3


COLLABORATIVE PROBLEM DEFINITION
COLLABORATIVE ACTIONPLANNING
EXPERIMENTAL INNOVATION
'PERMANENT' IMPLEMENTATION
OR
OR
IN-HOUSE DATA COLLECTION
FEEDBACK OF DATA
MONITORING & REVIEW
FEEDBACK OF RESULTS
PROBLEM RE-DEFINITION

which has been widely applied in many fields over recent years goes under the rather dramatic name of 'action research'. (Clark, 1972; Foster, 1972). Again, the process can be explained best by reference to Figure 3. This figure is rather more complex partly because the process is probably really more complex but also because it is rather better defined than the 'dissemination' model.

The fundamental differences between 'basic' research on the scientific method model and 'action research' can be seen clearly in this diagram. Firstly, defining researchable problems is a collaborative process involving practitioners and researchers. Secondly, the research is problem-focussed (although the data may be relevant to basic research issues); and thirdly, the process has application and evaluation built in through the experimental innovation stage.

Generally, action research is presented as a form of organizational problem-solving in which the external researcher plays the role of consultant. (The field of organization development is one which uses this model, for example, French and Bell, 1973). Clearly, however, this is not the only possibility and in the first project we shall describe, the process was what we shall call 'researcher-led'. By this we mean that the researchers were convinced that communication (and information) problems existed in the organizations concerned but that, in general, they would not be recognised as such by members of the organizations. This led the researchers to find settings in which to carry out action research and in doing so to demonstrate that their initial assumption about the existence of communication problems was correct.

This last point has a bearing on the difficulties likely to be faced by a researcher trying to negotiate entry to an organization to carry out action research. Some of the likely difficulties could be summarized as:

1. the problem postulated by the researcher may not exist, or it may be the surface effect of more deeply-rooted problems. If this is not detected before the work commences, the researcher may be faced with issues and difficulties for which

he is totally unprepared;

2. a variation of this is that the problem may not be recognised as a problem by the relevant management staff;

3. the researcher may believe that the source of the problem lies in the organization but management may think that it lies in the environment or in some controlling body;

4. the concern of the researcher, in presenting his perception of organizational problems, may be in research as an academic activity: it may be difficult to negotiate, therefore, if he fails to realize that the manager may be more concerned with organizational survival and growth;

5. acceptance of the researcher's ideas may be a management ploy, a move in a personal power-game, etc., rather than a genuine recognition that the problems outlined are real;

6. key individuals may seek to direct the research towards the justification of decisions already taken. (Based upon Foster, 1972).

Project INISS* was planned as an action research project from the beginning and three stages were envisaged:

- a period of data-collection through structured observation in social services departments, which would give the researchers insights into the day-to-day functioning of the organization, as well as allowing the development of research hypotheses (see Wilson and Streatfield, 1977);

- a stage of structured interviews to allow for the collection of a greater amount of data to test research hypotheses (see Wilson, Streatfield, and Mullings, 1979);

---

*These initials stood for two titles over the life of the project: 'information needs and information services in local authority social services departments' and 'information innovations in social services'.

- a period of experimental implementation of innovations in information services following upon the feedback of data from stages one and two. The choice of innovations would be made in consultation with departmental management and field staff (see Wilson and Streatfield, forthcoming).

In spite of a short gap in the duration of the project, which was the result in a change in funding organization after the interviewing phase (from British Library Research and Development to Department of Health and Social Security), the three phases detailed above were followed and the process followed the model in Figure 3, allowing for the fact that the whole project was 'research-led'.

It should be noted that the 'innovations' which were implemented were in no sense 'novel' to information workers: the novelty lay in the fact that virtually any information service was a novelty in the kinds of organizations in which the work was done. Initially, the funding agency (DHSS) was presented with a menu of almost thirty ideas for improving communication and information flow at all organizational levels. We were advised to focus our efforts upon the fieldwork level (that is, the level at which service is given to clients), and, as a result, the list was reduced to eight innovations. Following discussions with a number of departments (some of which had participated in the work from the beginning of the project) these innovations were eventually applied in a total of eighteen locations (area offices) in seven social services departments in three types of local authority: Shire County, Metropolitan District, and London Borough.

The innovations were:

- in-service training courses in communication and information handling (see Streatfield, Clark, and Hill, 1980);

- the development of local information collections relying upon organizational information and material from voluntary agencies, designed to be low-cost, easily-maintained collections;

- collections of monograph literature, supplied to area offices either by the local county library service or the department's own research library service (see Streatfield and Pritchard, forthcoming);

- an index to expertise;

- more effective journal provision;

- an abstract bulletin service (see Streatfield and Pritchard, op. cit.).

- assistance with the evaluation of a staff information bulletin; and

- the provision of background papers for in-service training courses.

Clearly, there is rather more work involved here than can be presented in detail in a paper such as this and we will concentrate on our general conclusions as they related to the usefulness of the action research mode. Anyone wishing to have more detailed information on any particular innovation is invited to contact the authors.

The conclusions of the innovation phase can be set out as follows:

- genuine collaboration with those directly affected by the innovation is absolutely essential. In one particular application it was clear that the area officer developed an 'instant enthusiasm' for organising information collections which was not recognised as a passing phase by the higher management or the researchers. Consequently, when her initial commitment to the idea fell away this was reflected in the attitudes of other members of the area staff who were used to these sudden bursts of enthusiasm. This innovation was markedly less successful in this office than it was in another department;

- genuine collaboration in defining the research problem is also important even with a 'researcher-led' programme. For example, the research team had originally intended to explore methods of providing information support for problem-solving groups such as internal working parties. This idea was transformed at the initiative of the host department (with the full agreement of the researchers) into broader support for in-service training courses);

- support from the local area management is essential. Again, in another test of the information collection innovation, the local area director was very enthusiastic (he had been one of our original observation subjects). He provided space, organized part-time manpower support and generally encouraged his staff to make use of the collection. In other innovations, also, where senior management were responsive to our ideas and where they believed benefits would flow to the organization, the task of introducing an innovation was made much easier;

- a user group which is apparently difficult to motivate to use information (social workers) will respond to the provision of information services at the workplace. Both the book collections and the information collections were used (in different ways, and with different degrees of intensity) in departments where long-distance use of central facilities had been virtually zero;

- to be useful the information provided must relate to the current problems or issues of interest to the recipient;

- it is possible to bring about sufficient organization change through action research to enable innovations to be adopted (or for other innovations to be adopted on the basis of lessons learned), after the experimental period. This can be achieved if the innovations are effectively monitored and evaluated and are seen by organizational staff and management to address problems of significance to the organization. Perhaps the most telling illustration of this was given to the authors after the conclusion of the project. A research officer reported that the abstract bulletin innovation had been virtually the only information service to survive drastic cuts in his budget. In fact, the service was now being provided to all area offices in the authority and the resources available were being used to buy more subscriptions to the bulletin (Social Work Information Bulletin) rather than to buy books or new journal subscriptions;

- one of the most effective ways in which to bring about an awareness of communication and information problems and a readiness to accept innovation is through training. In two authorities we organized training courses, in association with the departments' training sections. These have been described elsewhere (Streatfield, Wilson, and Clark, forthcoming), but the conclusions from the evaluation are worth repeating:

> 'During follow-up interviews it became clear that course participants were:
>
> - attempting to be more systematic in their preparation for and conduct of meetings;
>
> - more aware of the need to be conscientious in passing on information to their subordinates;
>
> - more aware of the information requirements of colleagues;
>
> - attempting to control their own information more effectively (e.g. by re-organizing and weeding filing systems);
>
> - active in promoting more effective information handling and use (e.g. by setting up team libraries, or reproducing and distributing course hand-outs); and
>
> - were aware of changes in other people's information related behaviour which could be attributed to the course'.

The second project described here came about in a different way and the role of the external researchers is likely to change between that of consultant, as usually envisaged in action research, and direct initiator of services. An element of speculation is inevitable in this case because this project has run only nine months of a four year programme and the final results of the continuing collaboration with participants cannot be anticipated.

In this example, the initiative to set up the Education Management Information Exchange (EMIE) as an experimental service came from the likely participants.

This came through the Society of Education Officers which is the professional body for senior administrators in local government education departments in England and Wales. Joint sponsorship for the research project in which EMIE is based is being provided by the Society, by the central government Department of Education and Science and by the National Foundation for Educational Research.

You will not be surprised to learn that we aren't yet ready to set out conclusions about this project! Instead we will show in figures 4 to 6 how we anticipate that the project will operate.

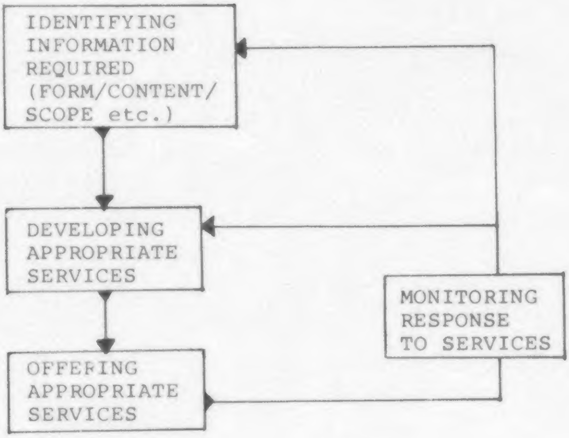
Figure 4 shows the various phases in our project programme. You will notice that we view the whole process as a continuous one throughout the life of the project. This diagram is pretty meaningless without some indication of what the monitoring phase is likely to involve and this we have shown in figure 5.

We suspect that the only service monitoring carried out by many information services is to record the number of requests received or items provided on loan. One might add the monitoring of user's comments but how many services record these systematically? The fact that the documentation part of our service will be supported by the IBM Stairs computer package holds out the possibility of our analysing the content of requests and looking at patterns in refining the search process for individual enquiries.

Turning our attention to potential service users I think it is worth stressing that we envisage the service as an information

FIGURE 4

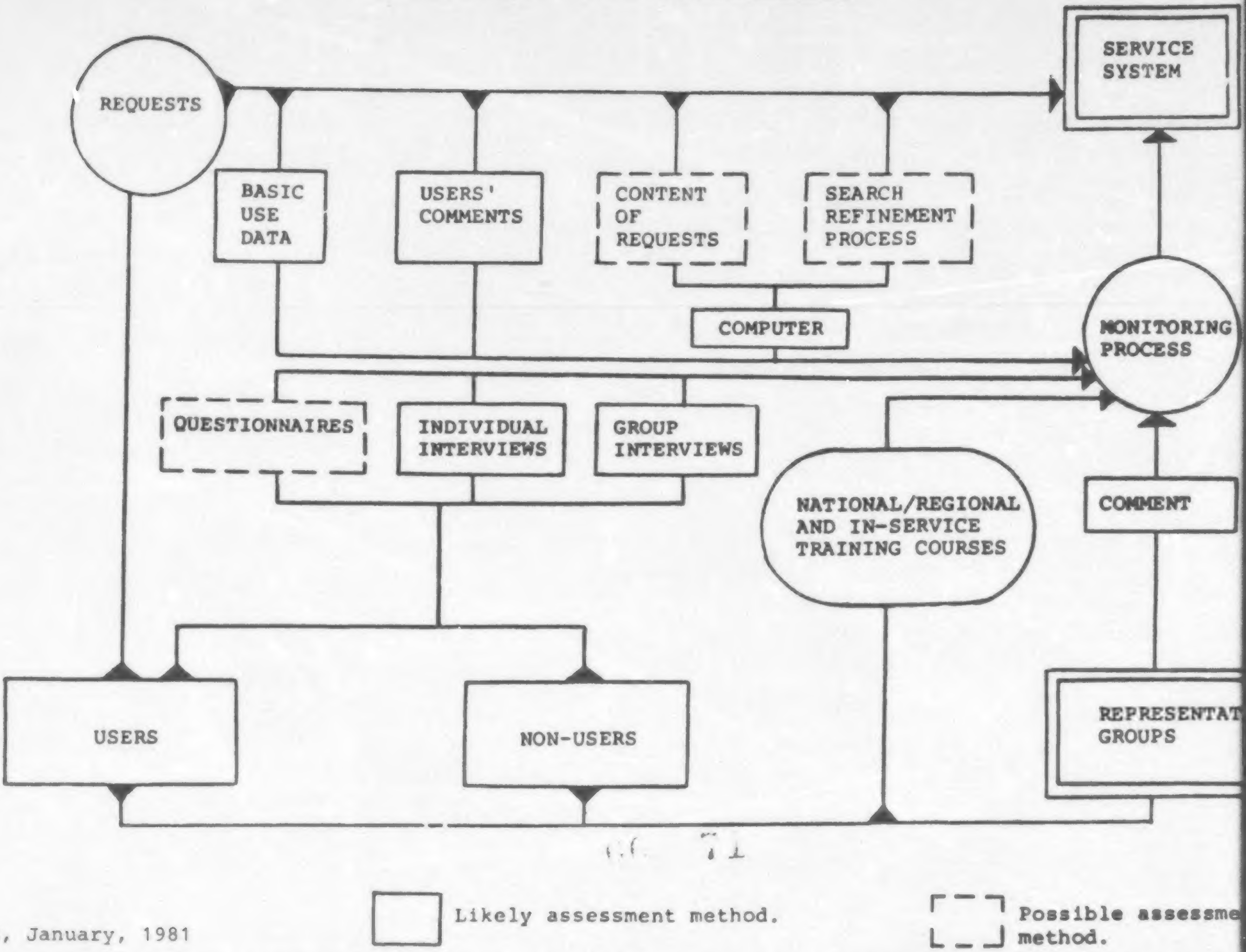
OPERATIONAL STAGES


IDENTIFYING INFORMATION REQUIRED (FORM/CONTENT/ SCOPE etc.)
DEVELOPING APPROPRIATE SERVICES
OFFERING APPROPRIATE SERVICES
MONITORING RESPONSE TO SERVICES

D R S, January, 1981

FIGURE 5.

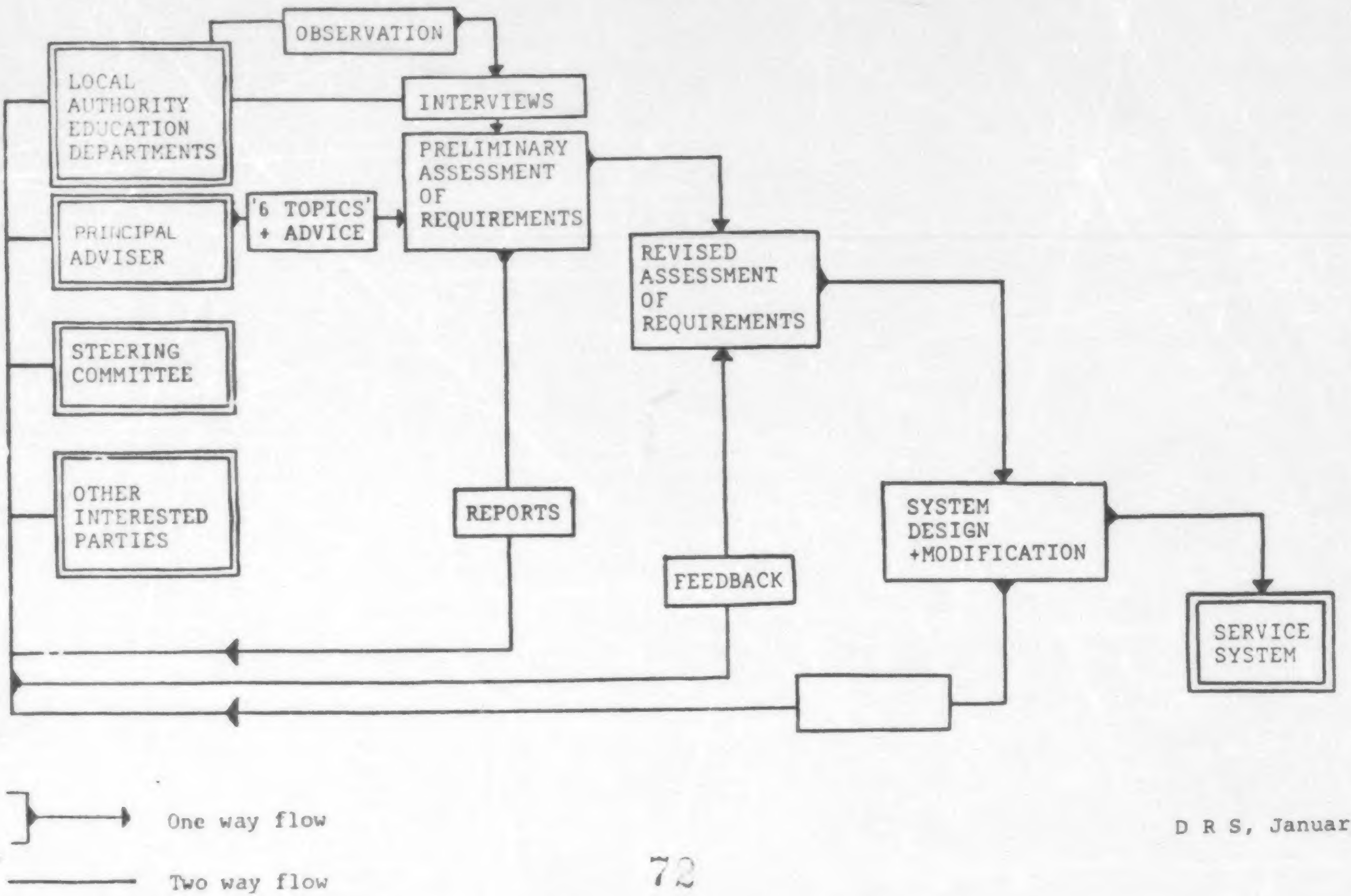
POSSIBLE APPROACH TO SERVICE MONITORING



D R S, January, 1981

IDENTIFYING INFORMATION REQUIREMENTS (PRE OPERATIONAL)

FIGURE 6.



D R S, January,1981

exchange. Not all the services offered are likely to work through our service centre and whatever services we do offer can only work if we ensure maximum participation by our users. For this reason we will be seeking practitioners' reactions to the problems encountered, desirable services and practical possibilities. We will do this through individual or group interviews and possibly through questionnaires. (For guidance in exploring these issues see Mullings, Francis, and Wilson, 1981). More significantly, we will seek active collaboration from practitioners by establishing a series of link-people in each local authority and we will involve practitioners in training courses looking at information in an organisational context. We envisage this activity as a very important part of the overall process because training courses of the kind we will offer should:

-enable practitioners to focus their attention on communication issues which they might otherwise take for granted;

-enable the project team to come into closer contact with the practical issues facing practitioners in their day-to-day work;

-encourage a much more equal exchange of problems and ideas between practitioners, information specialists and their advisers than could be achieved by an 'information officer as expert on information provision' approach;

-more specifically, this interchange should provide a chance to tell the project participants where they are going wrong and give them a chance to put it right.

At an informal level the project staff are already making heavy use of the Society of Education Officers as a point of access to education practitioners.

It will be obvious that much of what has been expounded in this account is based on Project INISS experience. This is even more true of the preliminary stage of the project which was concerned with finding appropriate areas of work for the information exchange service. Figure 6 shows the steps which we have already worked through in setting up the service. One weakness in the project programme is that the amount of observation and interviewing we are able to carry out is not large enough to ensure equal partnership between the researchers and practitioners at the next stage of involvement. The 'outsiders picture' of their work which the researchers presented to practitioners was a strong stimulus towards change in the social services project. In the current project the kind of picture which we can offer will be tentative. Fortunately, we will also be able to offer a comparative picture from the social services project and as the approach gains wider acceptance the results of comparative studies will increasingly become available to illuminate later work.

The next step will be to spend six months with link people and other staff from 30 of the 104 education authorities in England and Wales to test out some of our (and their) preliminary ideas for specific services. We are aiming to open the full scale service next March. I am looking forward to finding out what we will be doing then.

We have discussed some issues in applying action research methods to information science and looked at two examples. We will conclude with some general comments about the implications of this approach for information science generally and information service provision in particular.

In our view the 'university library model' of library service works well when the users are highly motivated to seek out the literature and when the literature is readily accessible. Similarly, the 'scientific research organisation' model of information service is appropriate when the users can readily define their subject interests and are interested in everything published on their specialisms and when bibliographical control of the literature is comprehensive, particularly if there is also a strong incentive to publish the results of research.

The proposition underlying our work and the one which led us to adopt an action research approach is that these two 'service-based' models of information provision work much less well when applied to other user groups, such as local government practitioners.

A cautionary note should be sounded here. The action research process is heavily dependant upon adequate evaluation by the participants in all the phases of the action, including evaluation of the evaluation process itself. This aspect of action research has so far received less attention than it merits. (Incidentally, we have published two papers on our own evaluation programme : Pritchard, Streatfield, and Wilson, 1979; Streatfield, Pritchard, and Wilson, 1981).

An action research approach to information service development may be difficult to 'sell' to sponsors because it is possible to go ahead and apply variations on the two models just outlined and be seen to be busy doing something, however peripheral to the main communication problems of the organization. However, an action research approach offers a number of advantages:

- it enables the information specialists to learn more about the real communication problems of potential service users;

- the problems will be seen in their organizational context;

- it should be possible to learn more about how practitioners do their work and about their perceptions of the part which information plays in that work;

- it provides a means for the information specialist's knowledge about information services and techniques to be explored from the point of view of their relevance to the problems in hand;

- it helps to raise awareness of the problems surrounding the communication of information both for practitioners and information specialists;

- it helps generate the conditions in which organizational change can be achieved.

In short, action research provides the means to redefine the relationship between the information specialist and the potential information user.

References

1. Clark, P.A. (1972) Action Research and Organizational Change, London : Harper and Row.

2. Foster, M. (1972) An introduction to the theory and practice of action research in work organizations, Human Relations 25, 529-556.

3. French, W.L. and Bell, C.H. (1973) Organization Development : Behavioral Science Interventions for Organization Improvement, Englewood Cliffs, N.J. Prentice Hall.

4. Mullings, C., Francis, G.M., and Wilson, T.D. (1981) A Manual for the Investigation of Local Government Information Needs, Sheffield, Postgraduate School of Librarianship and Information Science, U. of Sheffield, (British Library R+D Dept. Report 5585).

5. Pritchard, J., Streatfield, D.R., and Wilson, T.D. (1979) An illuminative programme for assessing information services, Aslib Proceedings 31, 245-50.

6. Streatfield, D.R. and Pritchard, J. (forthcoming) Will social workers read this? Published information in social services area offices. (To be published in Social Work Today).

7. Streatfield, D.R., Clark, D., and Hill, H. (1980). Training social services staff to make better use of information : an account of two practitioner-training courses and their implications, Aslib Proceedings 32, 134-44.

8. Streatfield, D.R., Pritchard, J., and Wilson, T.D. (1981) Assessment of an illuminative evaluation programme for information services, Aslib Proceedings 33, 67-70.

9. Streatfield, D.R., Wilson, T.D., and Clark, D. (forthcoming) A problem-based training approach to information for practitioners (To be published in Libri).

10. Wilson, T.D. and Streatfield, D.R. (1977) Information needs in local authority social services departments: an interim report on Project INISS, J. Documentation 33, 277-93.

11. Wilson, T.D., Streatfield, D.R., and Mullings, C. (1979) Information needs in local authority social services departments : a second report on Project INISS, J. Documentation 35, 120-36.

12. Wilson, T.D., and Streatfield, D.R. (forthcoming) Information innovations in social services departments: a third report on Project INISS (not yet placed).

EDUCATION OF THE END-USER IN COMPUTERIZED INFORMATION RETRIEVAL

Nancy Fjällbrant

Introduction

As long ago as in 1973, Lancaster pointed out that

> "a critical factor affecting the success or failure of an on-line information retrieval system is the effectiveness of the procedures employed to teach people how to use the facilities."[1]

Up to 1980, online education was primarily directed towards intermediaries, but in recent years an increasing amount of attention is being devoted to teaching end-users how to carry out online searching.

This paper will examine some recent work on facilitating end-user use of computerized information retrieval systems, and describe in detail end-user online instruction of engineering students at Chalmers University, Gothenburg, and the University of California at Santa Barbara.

End-user online interfaces - aids for the user

In recent years considerable effort has been devoted to making the computerized information system more 'user-friendly', that is, simpler to use in order to carry out efficient information retrieval. One approach has been the development of a computerized intermediary, or interface, between the end-user and the information retrieval system. An example of such a computerized search aid is the CONIT - COnnector for Network Information Transfer developed by Marcus at the Laboratory for Information and Decision Systems, M.I.T.[2,3] Users make their requests in a simple command language. These requests are translated by the computerized intermediary into the appropriate command language of the IR system to be used. The search aids include a search save process and a multiple database/system regeneration and repeat facility, an automated search technique developed from natural-language phrases via stemming and Boolean AND operation to automatically give a variety of searches based on the original phrase, aids to searching a number of individual databases, and a facility for automatic database selection. Preliminary results have been encouraging.

At Drexel University the IIDA system - Individualized Instruction for Data Access was developed based on a computer system which serves as an intermediary between a user and an online IR system. IIDA operates in two basic modes - as a teacher before a user begins to search and as an active assistant during a real search (diagnosing performance, commenting on errors and giving advice).[4] The IIDA system was tested on a group of industrial end-users in 1980.[5] The following comments were made:

> "The most significant conclusion is that there is a great interest in online searching by both scientists and engineers."
>
> "The study participants seemed to like using IIDA and online searching, which tends to support one of our hypotheses that end-users will like the idea of doing their own simple searches."

At the Information Retrieval Research Laboratory of the University of Illinois, Preece and Williams have worked on the development of a micro-computer based intelligent terminal for interface to a model search system.[6] User aids are envisaged as part of the model. A prototype The Searcher's Workbench (TSW) has been produced and the design of a production version considered.

Another approach to creating a user-friendly system has been worked on by Doszkocs at the National Institute of Health -

CITE Current Information Transfer in English - natural language query capability for the MEDLINE database.[7] With the aid of CITE, the search can be presented in English in free form - as sentences, phrases, lists of terms etc. No special syntax forms are imposed on the searcher. The CITE system translates the original query into MESH terms and allocates weights based on the frequency of occurrence in the lists of postings. Searching is performed and the retrieved documents are displayed in a ranked order, according to the satisfaction of the expressed search query as represented by the terms processed.

At Info-Mart, Bellamy has developed the 'Information Broker's Terminal System'[8] (BTS) which includes the following features:

1. The ability to review alternate search services (SOURCE) which provide access to the selected files (including comparisons of price, special system features, overlap with other available files, and other pertinent data about file characteristics).

2. A means for automatically translating a common set of keyboard command symbols into the several versions required by different search service retrieval systems.

3. A means to STORE search strategies, both primary and alternative, before going online to a search service, and for easily changing any segment of a strategy before reusing it at a later time.

4. A facility for block transmission of stored strategy segments with one command.

5. A means for automatically responding to messages from the search services, such as automatic logon, etc.

6. A facility for automatically recording (LOGGING) everything which transpires during a search session, including a record of any citations which are received online.

7. A facility for POST-search editing of any output taken online,

with most edit functions being performed automatically.

8. The ability to accommodate both new and experienced users with extensive HELP routines always available to anyone at any time.

End-user training and education

At the same time as the developments in user-system interfaces designed to reduce the number of variables that a user will have to cope with, attempts have been made to teach end-users how to carry out their own information retrieval on the existing systems and databases. One example of this is to be seen at the Library of Congress on the use of the SCORPIO system - Subject Content Oriented Retriever for Processing Information Online. SCORPIO was designed to meet Congressional information needs and the files contain information on all legislation for the present and past recent Congresses, abstracts of the printed Congressional Records citations to research materials or reports on public policy matters, full text of research reports written by the Congressional Research Service (CRS) for Congress, MARC records and descriptions of organizations useful as information sources. The data files vary in complexity from the simple bibliographic citation records to the information available on a given piece of legislation. SCORPIO was designed as a system for retrieving information in a logical way for a person unfamiliar with computer concepts. SCORPIO is an interactive time-sharing system manipulated by some 20-30 commands. Since 1978 regular classes in the use of SCORPIO have been provided for congressional staff. An average of 43 users per month have been trained.[9] These six hour introductory courses have enabled congressional staff to obtain a basic competancy in the use of the system. 72% of the trainees said that they still used the system up to one-and-a-half years after the initial training period, and 88% of these users used the system at least once a month.

Certain files of the SCORPIO system - MARC records - are now available for searching by the general public at the Library of Congress. People come in, sit down at a terminal and, with a minimum of instruction, search a large online catalogue. Most

81

of these end-users are very positive about this method of obtaining information.

The National Library of Medicine (NLM) has spent much effort on developing educational programmes and methods to teach the use of MEDLINE, for example the Computer Assisted Instructional (CAI) programme MEDLEARN.[10] It has been noted that there are an increasing number of end-users - medical practitioners - now attending the MEDLINE training sessions. A similar phenomenon has been observed at recent training sessions provided by the DIALOG information retrieval service. In a number of systems seminars, end-users have been in the majority.

A recent review article by Tedd describes a variety of teaching aids developed and used for education and training for online users. (Tedd, 1981).[11] Emphasis is on aids developed in the United Kingdom. Tedd describes audio and digital recording devices, Computer Assisted Instruction programmes for mainframe mini- or micro-computers, audiovisual aids and printed material.

Library User education programme at Chalmers University

Chalmers University Library had developed (from 1974 to 1980) an extensive user education programme[12] which could be roughly divided into the following stages:

1. Orientation for new users.

2. A 14-hour introduction to information retrieval for 3rd and 4th year engineering undergraduates.

3. Advanced courses in information retrieval (including computerized methods) for doctoral students - a 70 hour course.

4. Courses for practising engineers, laboratory assistants etc.

The 14 hour undergraduate course included a 20 minute, multimedia MEDIATRON programme, as an orientation to computerized information retrieval.[13] This was followed by short discussions on available online resources.

In 1980, Chalmers Library received a grant from the Office of the Chancellor of Swedish Universities, to investigate the possibility of teaching end-users how to carry out direct online searches on BYGGDOK. Over 300 undergraduates from the Schools of Civil and Mechanical Engineering have so far taken part in practical online training.[14] The BYGGDDOK online training will be described.

The BYGGDOK database

The BYGGDOK database has been developed and built up by the Swedish Institute of Building Documentation.[15] BYGGDOK is a bibliographic database containing some 30.000 references (with an annual growth rate of 5.000). The database covers the highly specialized field of building and construction with references dealing with the following subjects: architecture, building materials, building planning, civil engineering, construction technology, installation technology and urban planning.

Input to the database is mainly Scandinavian material, in the form of journal articles, reports, product information standards etc. The abstracts are written in the language of the country of origin (mainly Swedish, Danish and Norwegian). Each reference is described with Swedish keywords. The database can, however, be used by persons with no knowledge of Swedish, by means of a Reduced UDC code system developed by the Swedish Institute of Building Documentation. BYGGDOK is available online to users throughout Scandinavia via SCANNET, the Nordic data network and within the CIBDOC-system via EURONET. The database is widely used by practising engineers and architects.

Goals and objectives for the BYGGDOK online instruction

The main goal for the undergraduate online programme was to enable the student to carry out online information searches within his/her subject field on the BYGGDOK database, as, and when, required in connection with information needs.

Specific objectives were described as follows - after completing the course the end-user should:

1. Be aware of different methods for information retrieval: manual, batch processing, SDI and interactive on-line

2. Be aware of different kinds of information searches: current awareness, retrospective, factual

3. Be able to carry out logging-on procedures for a network/database

4. Be able to use a teletype or VDU terminal

5. Be able to use the basic commands and features of the BYGG-DOK retrieval system

6. Be familiar with a typical unit record

7. Be able to express a search topic in suitable terms and parameters

8. Be able to develop a search strategy in terms of Boolean logic

9. Be able to evaluate the results obtained

Teaching methods chosen and organization of course

There is a wide variety of methods and media available for on-line teaching. A selection of methods was used - lectures, small group seminars and demonstrations, audio-visual MEDIATRON demonstrations, printed manuals and checklists, and direct 'hands-on' training. The organization of the undergraduate course is shown in Fig. 1.

The introductory course lecture was used to present a frame of reference, with emphasis on the patterns of communication and the relationship between computerized information retrieval and manual methods. Different types of information search were presented.

The main part of the course was based on small group methods -

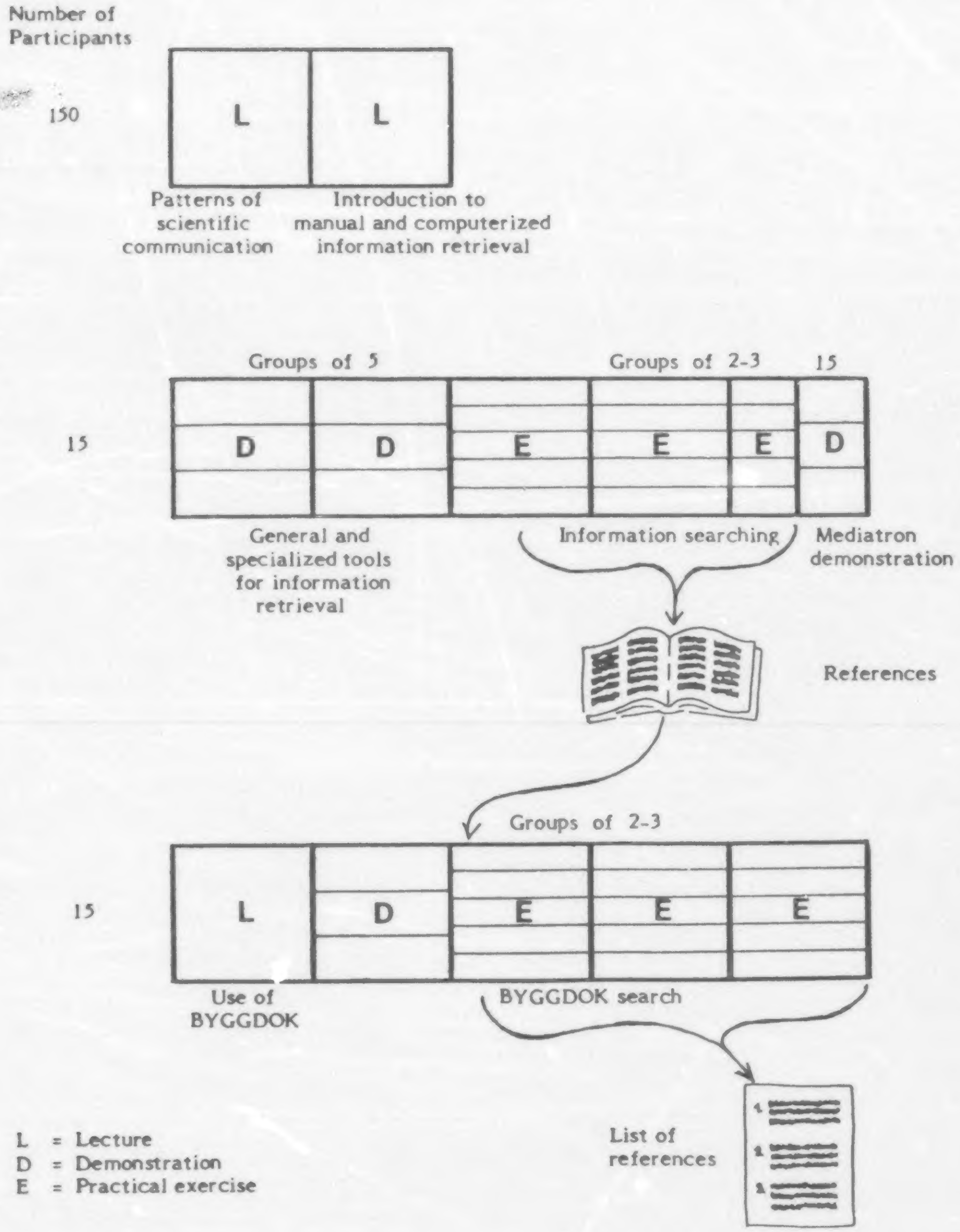


Fig. 1. Diagram of introductory course in information retrieval at Chalmers University of Technology

laboratory sessions with demonstrations and practical searching on a topic of the student's own choice. The first laboratory session (5 hours) was mainly devoted to manual information searching, using a variety of tools. The last hour was used for an orientation on computerized retrieval using audio-visual multimedia MEDIATRON programmes. The BYGGDOK database was described and demonstrated. Students were then given a short manual which introduced computerized information retrieval, described the BYGGDOK database, logging-on techniques, the search command language and search techniques. They were told that at the next session, in a week's time, they would carry out their own hands-on search.

The second laboratory session started with a short repetition, including a MEDIATRON demonstration programme of BYGGDOK. Students then prepared their searches, planning strategies with the aid of search-term frequency lists. They usually worked in groups of 2 or 3 persons. The group then carried out their search on a teletype terminal. They had a short BYGGDOK 'check list' of commands placed as an aid, immediately adjacent to the terminal.

Evaluation

Evaluation of the online learning was carried out in a number of ways:

1. Students were asked to complete a questionnaire about:

   - overall gain from the course
   - the teaching material used
   - the practical terminal sessions
   - future use of the BYGGDOK database and of online searching
   - the teaching methods used

2. Direct observation of how the end-users managed to log on and off and carry out a search

3. Comparative search analysis with regard to

   - the process (search effort)

   - the product (in terms of recall and precision)

A number of end-user searches were repeated by a trained BYGG-DOK intermediary. Searchers' efforts were compared with respect to:

Number of commands used

Different types of commands used

Numbers of search descriptors

Sets viewed

Search modification

Errors made

Connect time

The choice of these parameters was based on measures for discrimination among users with different levels and experience of online searching described by Fenichel in 1979.[16] The searches were also compared with respect to recall and precision.

Questionnaire responses

Students were asked to state their opinions to 'overall gain' from the course. Responses are shown in Table 1. For comparative purposes similar responses (for civil engineering undergraduates) to a course on manual methods for information retrieval are shown.

Table I. 'Overall gain' from courses in information retrieval

'Overall Gain'	'Online' course 1980 %	Manual course 1975 %
Nothing	0	0
Little	0	3
Satisfactory	16	24
Good	67	58
Very good	17	15

Students were asked to give their opinions on the new teaching material produced. Responses are shown in Table II.

Table II. Opinions about teaching and media

	BYGGDOK manual	MEDIATRON demonstration	Check-list
Very good	14	9	19
Good	73	47	61
Satisfactory	12	36	20
Hardly satisfactory	0	8	0
Unsatisfactory	0	0	0

Students were asked to evaluate the practical terminal sessions. See Table III.

Table III. Practical terminal sessions

Terminal sessions			
attitudes to		difficulty	
Very good	23%	Very easy	23%
Good	70%	Easy	63%
Satisfactory	6%	'About right'	14%

Students were asked to state their attitudes to this kind of training in computerized information retrieval. Their opinions are given in Table IV. For comparison a group who had merely received orientation are included.

Table IV. Do you think this is a good way for teaching about online information retrieval?

	1980 (includes practical searching)	1978 orientation only
Very good way	61%	10%
Good way	38%	66%
No opinion	1%	9%
Not particularly good	0%	7%
Not a good way	0%	0%

Students were asked if they thought they would use the BYGGDOK database at some future time. 55 % thought they would use the database, 3 % thought it too expensive and 42 % were not sure. 84 % stated that, if they had access to a terminal, they would like to carry out their own online searching in the future, whereas 17 % were not sure about this. 80 % of the students said that they preferred computerized information retrieval to manual methods. Reasons given were speed, convenience and flexibility. A number of students pointed out that it was possible to combine more search terms in computerized searching than by use of the manual tools. One particularly interesting fact that emerged from the evaluation, was that students were themselves evaluating the different approaches of manual and computerized information retrieval.

Observations

All the students (some 300) who have so far taken part in online training sessions, have managed to carry out the various logging-on procedures, and carry out the terminal searching. In the first sessions the main difficulty was finding the ON/OFF switch on the terminals. (This was easily remedied by writing ON and OFF marks in large red letters on the appropriate parts of the machine!) Most of the students worked in small groups, and this appeared more satisfactory than individual terminal work. In cases of difficulty the group could often work out a joint solution, whereas individual searchers tended to ask the assistant in charge of the session for advice.

There was no doubt that the students enjoyed the terminal sessions. Often it was difficult to stop students searching! This sometimes led to grumbles amongst those waiting. We increased the number of terminals from 2 to 4 in order to limit waiting time. A considerable number of students have come back to the library and asked to carry out an online search in connection with some other project.

Comparative search analysis

A number of end-user searches (11) was compared with similar searches carried out by a trained BYGGDOK intermediary. The latter, not only had extensive experience of searching, but also of input and construction of BYGGDOK. She was conversant with the use of search technique based on the special reduced UDC developed at the Swedish Institute of Building Documentation for this purpose. At the same time, the control searches were carried out in the absence of an end-user, so that the interactive relevance judgement of the user was missing. The intermediary said that, as a non-engineer, she experienced some difficulty in relevance judgements. The student group was highly motivated, as they were searching for information needed for their research projects. With respect to external conditions, variables such as type of terminal, search tools, database and search command language were controlled and identical. System response varied to a small extent, from search to search. Comparison of search effort between end-users and intermediary are shown in Table V.

From Table V, it can be seen that the end-users expanded a greater effort on their searching than the intermediary. The latter carried out short, economical searches. The end-users showed considerable determination and flexibility in their attempts to obtain relevant references from the BYGGDOK database, with repeated restarts and the use of a variety of descriptors. Their inexperience can be seen in the errors (4 per search) and in the dominance of use of non-truncated descriptors, with respect to truncated terms. The experienced intermediary made greater use of truncated descriptors and searching by means of reduced UDC terms. (Students were not taught the use of the reduced UDC search method). Connect time was greater for end-users than for the experienced intermediary.

Table V. Comparison of search effort - BYGGDOK searches

Search effort variable	End-users *	Intermediary*
Total no. of commands used	26	8
Different types of commands	8	5
Total number of descriptors searched	11	4
Number of truncated descriptors searched	4	2
Number of non-truncated descriptors searched	7	2
UDC-descriptors	0	1
Unique descriptors	8	4
Sets viewed	3	1
Errors/search	4	0
Connect time	<	>

*Average/search

Results of end-user and intermediary searches were compared with regard to recall and precision. These are defined as follows:

$$\text{Recall} = \frac{\text{No. of relevant references retrieved}}{\text{Total no. of relevant references in the database}}$$

$$\text{Precision} = \frac{\text{Relevant references retrieved}}{\text{Total no. of references retrieved}}$$

The search results are seen in Table VI.

Table VI. Comparison of search results

Search product variables	End-users	Intermediary
Recall	62%	80%
Precision	46%	76%

Recall is surprisingly high for the end-user group. This may be connected with their search motivation and aforementioned determination to obtain references. Precision was considerably lower for the end-user group than for the intermediary. This factor is perhaps of less importance in the use of a small database than with a larger one.

Searching by undergraduate engineering students at UC Santa Barbara

In 1980 and '81 third year mechanical engineering students were given training in online information retrieval as part of a course in engineering design, at the University of California Santa Barbara (UCSB). The concept and methods for online information retrieval were presented in lecture form by a member of the academic staff. Students were provided with a handout on the use of the DIALOG information retrieval service giving information on:

logging on and off techniques
basic commands
Boolean operators
search commands
output control
detailed information, including a sample record, on the use of File 14.
a sample search from NTIS

Students were then given search assignments to perform with the aid of the DIALOG classroom search facility (DIALOG provides a special educational training facility with access to a limited number of files). Students could either choose to search on one of six given topics or on a subject of their own choosing. The results of 32 of these searches carried out by some 90 students were analysed with respect to search effort and search product in a similar way to the searches on BYGGDOK described previously. Searches on the set topics were carried out by a trained intermediary. Observation showed that many students experienced difficulties in logging-on and in carrying out their searches.

The analysis of 7 searches on the standard set topics, with regard to search effort is shown in Table VII.

Table VII. Comparison of search effort - set search topics using DIALOG

Search effort variable	End-users*	Intermediary*
Total no. of commands used	23	48
Different types of commands	7	11
Total numbers of descriptors searched	10	20
Number of truncated descriptors searched	>1	12
Number of non-truncated descriptors searched	9	8
Unique descriptors	5	9
Sets viewed	3	3
Errors/search	4	2
Number of databases used	2	3
Connect time	>	<

*Average/search

The 32 searches examined showed that very few of the students had made use of truncated descriptors, and relatively few had utilized the 'expand' command. It seemed that most of the mechanical engineering undergraduates decided on a search strategy, transmitted this to the computer and made few interactive modifications to the original strategy. They tended to either obtain a very small final set of references (about 10) or a very large (300-500). Few attempts were made to broaden or narrow these search results. A number of searchers, (11 out of 32), used more than one database.

With regard to the search product, recall for 10 searches performed on the given topics was very low - 17 %. For searches on a topic of their own interest, the mechanical engineering undergraduates had a higher recall - 36 %. Searches with a high recall showed low precision.

Discussion and conclusions

Most of the teaching/learning activities on computerized IR for end-users has been directed towards orientation and promotion -

by means of demonstration searches and lectures. There is, however, evidence to show that new users can learn to perform searches after a short period of training. Engineering and science undergraduates are, today, in many cases, competent in the use of terminals for online computations and simulations. It seemed, therefore, reasonable to test the hypothesis that end-users could be trained to carry out their own interactive online searches.

The study at Chalmers showed that 3rd and 4th year undergraduate engineering students could be taught to use a relatively small database - BYGGDOK - for information retrieval. The students achieved high recall values; these high recall values are probably due to the relatively small amount of relevant material present and to the students' high level of motivation. The students expressed very positive attitudes towards online information retrieval - 84 % said that, if they had access to a terminal, they would like to carry out their own online searching in the future. The online training seems to have increased the students' awareness of methods of information retrieval in general.

The positive experience from this experiment led to the introduction of hands on online training for a group of 200 mechanical engineering undergraduates, in 1981, on MechEn - a database produced at the Royal Institute of Technology, Stockholm. A second group of civil engineering undergraduates, and a smaller group of postgraduates have received training in the use of BYGGDOK.

The analysis of the results of the first-year mechanical engineering students at UCSB in the use of DIALOG showed much poorer performance. This is hardly surprising, in view of the much larger system involved - with choice of many databases - and the relatively short period of instruction. Yet all the groups, with one exception, managed to carry out some sort of search. The results have to be compared with the kind of manual searches which are carried out by first year engineering students! Students from the 1980 class have used the IR search techniques they learned that year, in project work in 1981. The standard of searching was reported to be much better than for the first year students. The importance of motivation for end-user search success cannot be overemphasized.

A group of electrical engineering graduate students at UCSB showed considerable aptitude for carrying out their own searches on the DIALOG system, after a short period of training. It is hoped to extend this programme in the near future.

It is often argued that intermediary performed searches are more cost effective than those of end-users. Cost effectiveness is sometimes expressed in terms of unit cost where

$$\text{Unit cost} = \frac{\text{Connect time cost}}{\text{No. of relevant references retrieved}}$$

This measure of cost effectiveness is, however, limited. In the world of the practising engineer or scientist, time is precious, and a truer measure of cost effectiveness would be

$$\text{Cost effectiveness} = \frac{\text{Cost of time used by engineer}}{\text{No. of relevant references}}$$

or in the intermediary performed search

$$\text{Cost effectiveness} = \frac{\begin{array}{l}\text{Cost of time used by engineer}\\ \text{+cost of time used by intermediary}\end{array}}{\text{No. of relevant references}}$$

It is planned to examine the cost effectiveness of end user searching in this way.

In conclusion, it has been shown that it is possible to train end-users to carry out reasonably effective searches on one database and system. It is more difficult to do this with increasing system size and complexity. Nevertheless the results obtained with end-users on the DIALOG system are sufficiently promising to stimulate further work. With user-friendly interfaces, it seems reasonable to suppose that end-users will be increasingly able to carry out their own information searches in the future. The role of the intermediary, in turn, might well develop from that of individual searcher, to that of adviser and consultant, with specialized knowledge for transdisciplinary or exhaustive searches.

> "If you give a man a fish
> He will have one meal
> If you teach him to fish
> He will be fed for life"
>
> Old Chinese Proverb.

References

1. LANCASTER, F.W. & FAYEN, E.C. Information Retrieval On-Line. Los Angeles, Melville, 1973.

2. MARCUS, R.S. & REINTJES, J.F. Experiments and analysis on a computer interface to an information retrieval network. M.I.T. Laboratory for information and Decision Systems Report LIDS-R-900, April 1979.

3. MARCUS, R.S. Search aids in a retrieval network. In: Communicating Information. Proceedings of the 43rd ASIS Annual Meeting, Annaheim, California, October 5-10, 1980, pp. 394-396.

4. MEADOW, C.T. The computer as a search intermediary. Online Vol. 3 (1979) no. 3, pp. 54-59.

5. LANDSBERG, M.K., LAWRENCE, B., LORENZ, P.A. & MEADOW, C.T. A joint industrial-academic experiment: an evaluation of the IIDA system. In: Proceedings of the 43rd ASIS Annual Meeting, Annaheim, California, October 5-10, 1980, pp. 406-408.

6. PREECE, S.E. & WILLIAMS, M.E. Software for the searcher's workbench. Ibid, pp. 403-405.

7. DOSZKOCS, T.E. & RAPP, B.A. Searching MEDLINE in English: a prototype user interface with natural language query, ranked output and relevance feedback. In: Information Choices and Policies. Proceedings of the 42nd ASIS Annual Meeting, Minneapolis, Minnesota, October 14-18, 1979.

8. FJÄLLBRANT, N. Reserapport för TEMADOC. Studiebesök i USA, våren 1981, DFI reserapport, Stockholm, 1981.

9. GRIFFITH, J.C. & NORTON, M.P. Training congressional staff to search the SCORPIO online system. In: Proceedings of the 43rd ASIS Annual Meeting, Annaheim, California, October 5-10, 1980, pp. 348-350.

10. EISENBERG, L.J., STANDING, R.A., TIDBALL, C.S. & LEITER, J. *MEDLEARN* A computer-assisted instruction (CAI) program for MEDLARS*. Bull. Med. Libr. Assoc. Vol. 66 (1978) no.1, pp. 6-13.

11. TEDD, L.A. Teaching aids developed and used for education and training for online searching. Online Review. Vol. 5

(1981) no. 3, pp. 205-216.

12. FJÄLLBRANT, N. *The Development of a Programme of User Education at Chalmers University of Technology Library*. Ph.D. thesis. Guildford, University of Surrey, 1976.

13. PRATT, G. & VICKERY, A. The development of multi-media teaching aids for users of computer-based information retrieval systems. *Program*, Vol. 11 (1977) no. 1, pp. 10-15.

⑭ FJÄLLBRANT, N., KIHLÉN, E. & MALMGREN, M. End-user training in the use of a small Swedish database. Paper presented at the ALA 100th Annual Conference, San Francisco, June 26 - July 2, 1981.

15. ABRAHAMSSON, S., BACKLUND, F & ERESUND, B. Costs of building up small databases in highly specialized fields or in restricted language areas. Paper presented at the 1980 Annual Meeting of FID "Information for Industry", 15-17 August, Copenhagen, 1980.

16. FENICHEL, C.H. Online searching: Measures that discriminate among users with different types of experience. *Journal of the American Society for Information Science*, Vol. 32 (1981) no. 1. pp. 23-32.

THE DISSEMINATION AND USE OF SCIENTIFIC AND PROFESSIONAL INFORMATION IN MEDICINE

Tuula H. Laaksovirta

Let me begin by presenting the picture of the philosophical thoughts of my research.

Figure 1 is supposed to make clear to you my cognitive way of thinking. Quite often in information science we study the information needs of a person. We suppose that people try more or less actively themselves to solve their information problems. My starting point has been the opposite. I have been interested in mediation mechanisms and diffusion and dissemination processes from the point of view of a society.

The purpose of my study on dissemination and utilization of medical information is to critically examine the theoretical frameworks of dissemination and utilization studies in general.

To study dissemination also on a concrete level, I'll give a summary of the results of two empirical studies, which I have made in the years 1978 and 1979.

Terms and concepts which I use arise from the sociology of science or from science studies. This is because I have tried to adapt the thinking of science studies to information science. Usually the researchers know and use the results of information science like Bernal and de Solla Price, but information scientists have not used science studies as much.

The number of studies of the dissemination and utilization of knowledge has greatly increased during the last twenty years. In 1954 only 50 studies a year were published in English. Ten

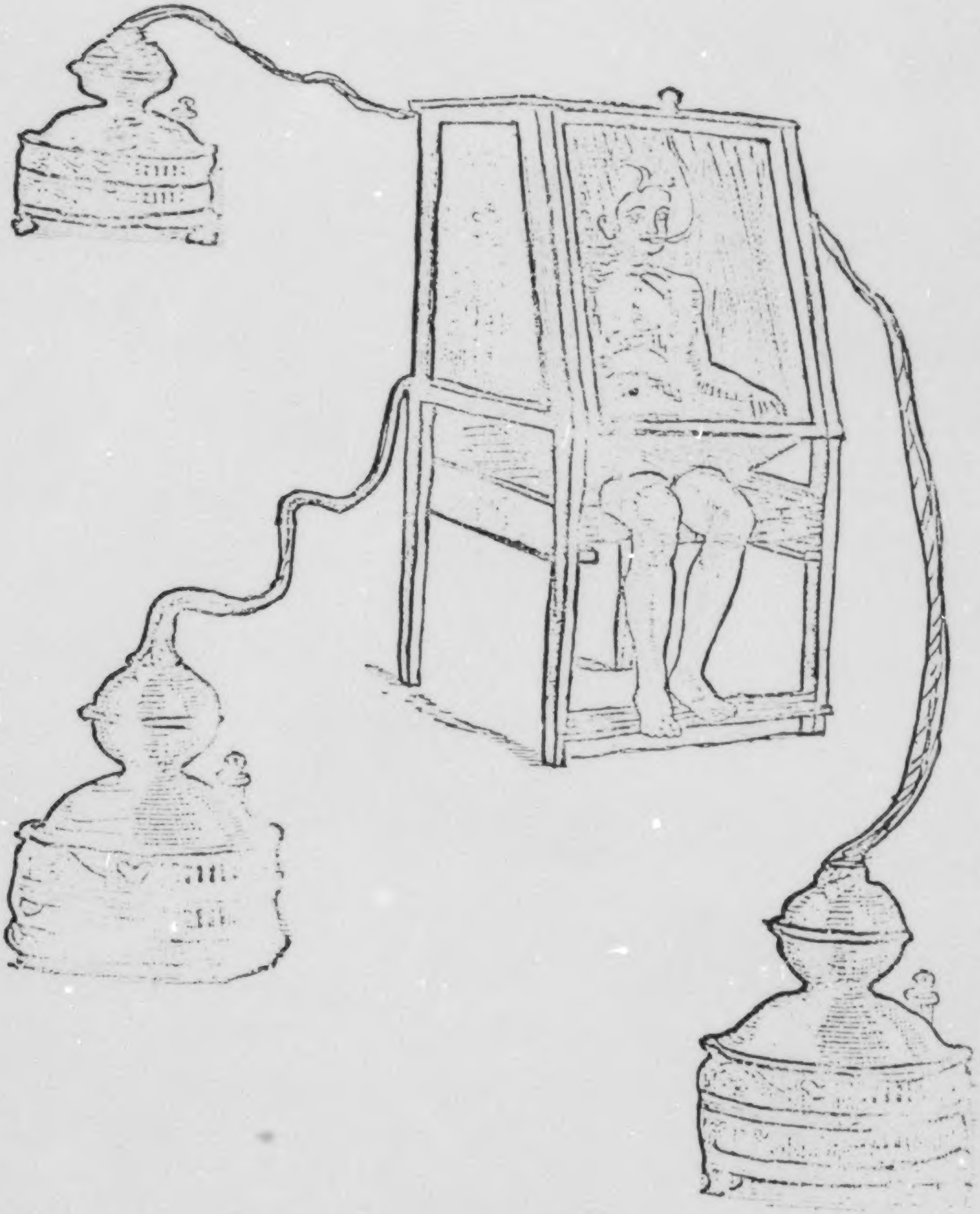


Figure 1

years later the number was over 500 and at the beginning of the 70s at least 10.000 studies were published in this field. (Havelock et al. 1973, p. 11-1)

In order to get an overview of the studies and their frameworks, Havelock and his research group (1973) have made a research of 10.000 studies. They picked out 4.000 studies for analysis. I'll now describe and comment on the main results of their study. (1973).

They found out that three main types of frameworks were used in dissemination studies. They called the first type of framework 'The Research, Development and Diffusion Perspective'. (Havelock et al, 1973, p. 11-6) (See Figure 2)

The main features of this kind of studies are that the dissemination of information is understood as a rational process or a simple process: from basic research to applied research - disseminating and use. It is supposed that there is this kind of division of labour in science and that high investment in information services pay off in quality and quantity.

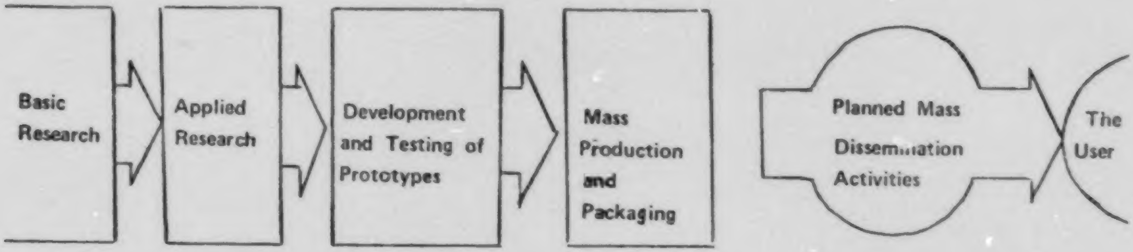
We all know that this model is too idealistic and that the reality is not at all as rational.

Havelock and others called the second type of framework in these studies 'The Social Interaction Perspective'. (1973, p. 11-8) (See Figure 3)

The central idea is that dissemination depends on personal relationships. Group memberships and interactions are important. The same applies to social structure as well as to power and influence structures. Opinion leadership structures are very important. This model is quite often used in information sciences when we study invisible colleague systems and the networks of researchers, for instance science citation index method is often used in this connection.

The third type of framework used in these studies after Havelock is called 'The Problem Solver Perspective'. (1973, p. 22-12) (See Figure 4)

RESEARCH, DEVELOPMENT & DIFFUSION PERSPECTIVES


Basic Research
Applied Research
Development and Testing of Prototypes
Mass Production and Packaging
Planned Mass Dissemination Activities
The User

(Havelock et al, 1973)

Figure 2

THE SOCIAL INTERACTION PERSPECTIVE

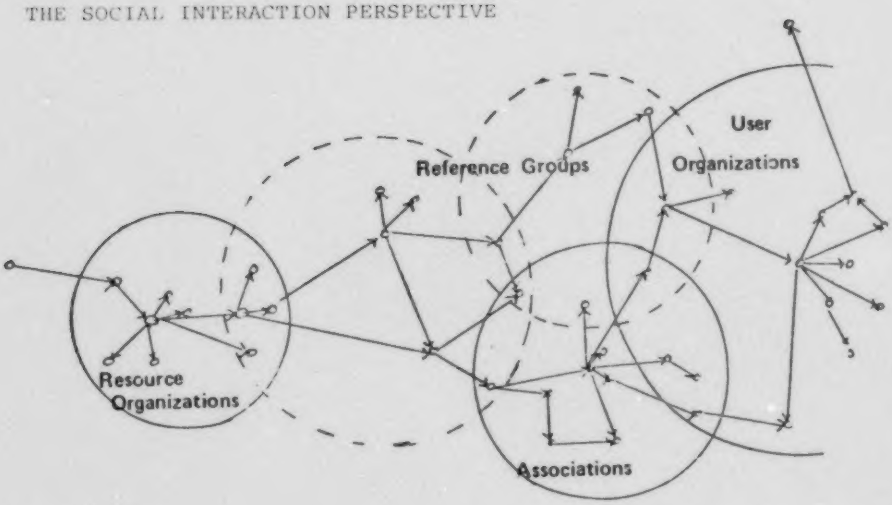


Figure 3 (Havelock et al, 1973)

THE PROBLEM SOLVER PERSPECTIVE

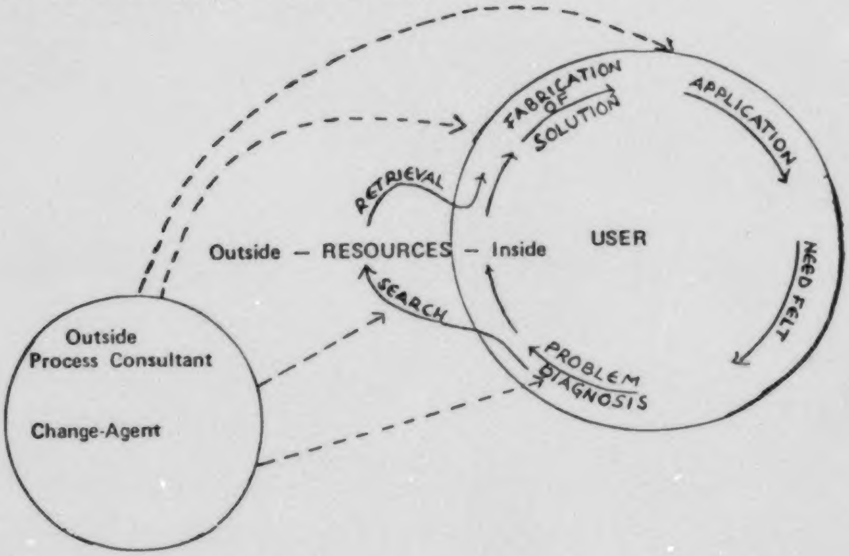


Figure 4 (Havelock et al, 1973)

Its main features are that obtaining knowledge is but one part of problem solving. The user's need is a paramount consideration. The phases of problem solving are: realization of the problem; definition of the problem; the search for relevant resources, the fabrication of the solution, and the application of information to solve the problem. The user must come to a solution himself; the outsider, for instance the librarian, is only a catalyst and consultant.

After having presented these three frameworks, Havelock suggests his own new framework. (1973, p. 11-16) (See Figure 5)

He calls this new framework 'The Linkage Process Model' and suggests that this model, or framework, should be used in dissemination and utilization studies. As the authors put it, the resource system must recapitulate or adequately simulate the user's problem-solving process. The user must be able to understand (and simulate) the research, development and evaluation process employed by the resource system in the fabrication of solutions. Resources and the user must provide reciprocal feedback. In the end they supposed that successful linkage experiences build channels for efficient dissemination. In this framework which Havelock and others presented to be used, it was assumed that the user for instance in information services gets the right answers to the right questions. But is it correct to assume this?

In my opinion the dissemination of information has to be seen as an ongoing process between science and society and not only between the user and resources for problem solving.

The other weakness of the last model which they recommended for our use in studies, is that dissemination and utilization are seen as a way of asking questions and getting answers. But is it right to assume that the user can ask questions at all? Asking questions requires scientific knowledge of for instance medical problems. In fact most medical information does not reach the user because he asks questions. It reaches him through all kinds of publications which he does not ask at all and through all kinds of persons which he happens to meet during the day.

THE LINKAGE PROCESS

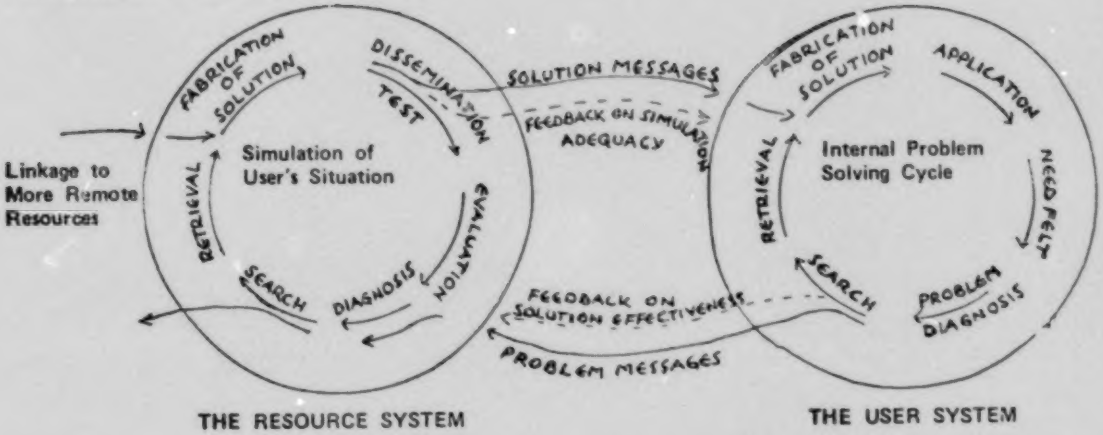


Figure 5

(Havelock et al, 1973)

There are still two other reasons why I criticize the so called linkage-process-model. It does not allow for conflicting interests in the dissemination of information. This is to say that with this model it is not possible to handle conflicting interests in the dissemination and utilization of research results which is anyway everyday life in science. We could say that it is very important to know how information is hindered to disseminate. Reasons for prohibiting the dissemination of information can be, for instance, making profit by research results as in medical industry, or making war as in technical sciences, or gaining personal honour as the story of discovering DNA tells.

In order to study dissemination and utilization questions on the empirical level, I'll summarize now two empirical studies which I have made of dissemination of information in medicine.

One of them was a questionnaire to 1.000 physicians; 811 physicians replied. The other one was a content analysis of Finnish medical journals from the 60s and 70s.

The first one tries to answer the question of adequacy of theoretical frameworks presented above and of the process of disseminating information in medicine. The other one tries to answer the question on the structure of the medical journals of the 60s and 70s. (See Figure 6)

The results of the first study showed that the most important institutions for the dissemination of medical information to physicians were the Medical Association, the pharmaceutical industry and the professional community in general.

In my study I divided all physicians into three groups: hospital physicians, health centre physicians, and university physicians. The institutions which are closest to science, that is to say university libraries and scientific societies, are less important institutions to health centre physicians. Nonspecialized health centre physicians usually don't hold doctoral degrees. Particularly this group is very dependent on the pharmaceutical industry for getting information.

THE INSTITUTIONS OF THE DISSEMINATION OF MEDICAL INFORMATION
(RATHER OR VERY OFTEN)

DISSEMINATION INSTITUTIONS (%)

Hospital physicians	Health center physicians	University physicians
TRADE UNION 91	TRADE UNION 93	TRADE UNION 93
PROFESSIONAL COMMUNITY 80 72	PHARMACEUTICAL INDUSTRY 56	PROFESSIONAL COMMUNITY 70 69
SCIENTIFIC LIBRARIES INFORMATION SERVICES 56 41 24	PROFESSIONAL COMMUNITY 57 50	SCIENTIFIC LIBRARIES INFORMATION SERVICES 62 58 45
SCIENTIFIC SOCIETIES 32	SCIENTIFIC LIBRARIES INFORMATION SERVICES 45 24 24	SCIENTIFIC SOCIETIES 60
PHARMACEUTICAL INDUSTRY 35	SCIENTIFIC SOCIETIES 10	PHARMACEUTICAL INDUSTRY 17

Figure 6

In information science we study mainly the dissemination process within libraries and information service systems and we study mostly the distribution of research results among researchers and scientists. My results show that we ought to study all groups which use scientific results. The same results show that we ought to study all systems which disseminate scientific information. (See Figure 7)

The study also showed that the most important communication medium of physicians were the professional journals. We can say that all categories of physicians regularly read the journals. As can be seen from the figure, the next best sources of information dissemination are the working place and colleagues. The implication of this seems to be that scientific media are not at all that important in keeping physicians up-to-date on medical information.

The objective conditions of physicians are not only reflected in their sources of information and in the selection of the material they use but also in reading methods, in the time spent by them on receiving information. (See Figures 8a and 8b)

On the average, physicians use 2.4 working hours a week to keep themselves informed of medical knowledge. In addition they use 4 hours a week of their leisure time receiving medical information. But again it must be said that the differences between the different groups of physicians are considerable.

University physicians are also more active in the reading of medical books than the others, and perhaps surprisingly, all physicians read as much other kinds of literature. (See Figures 9a and 9b)

Although all journals play an important role in the dissemination of information, there are big differences in how different types of journals are read. (See Figure 10)

The two at the top of the list are the Finnish Medical Association Journal and the Finnish Scientific Journal. These two journals are also very important media for advertising. The other

THE MEDIA OF THE DISSEMINATION OF MEDICAL INFORMATION

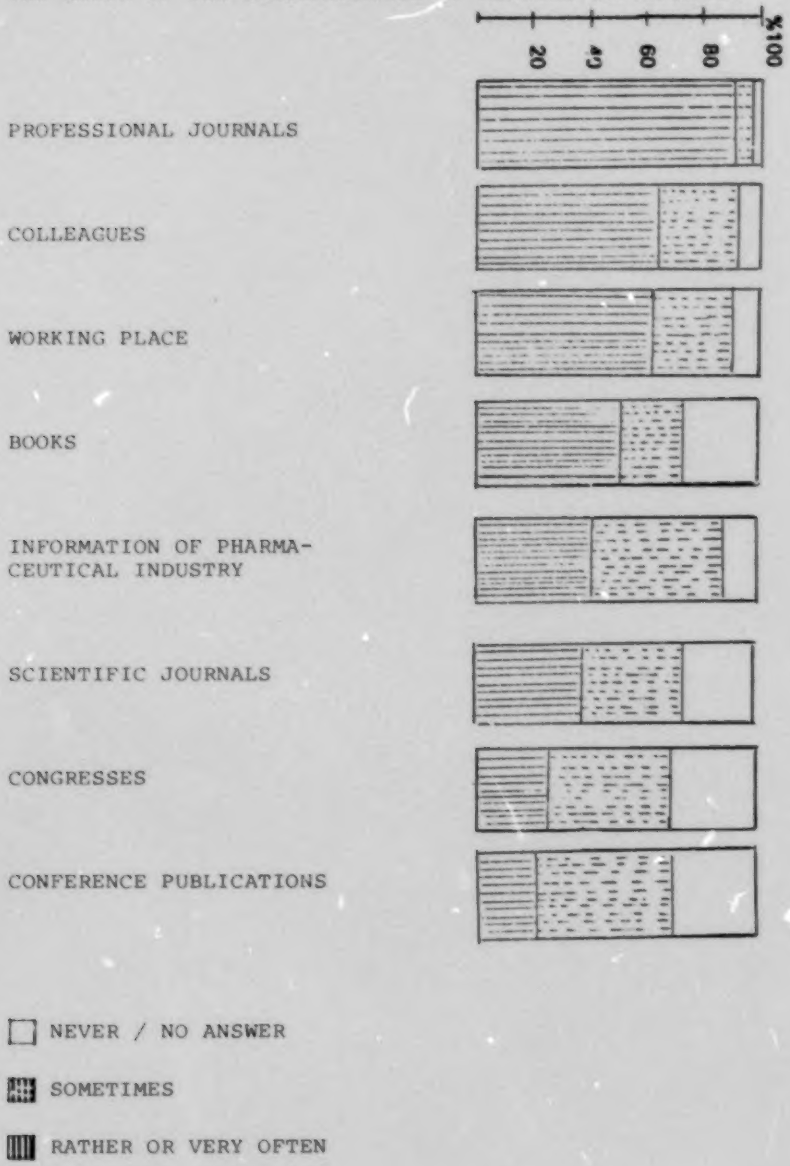


Figure 7

THE TIME SPENT ON FOLLOWING MEDICAL INFORMATION;
WORKING-TIME

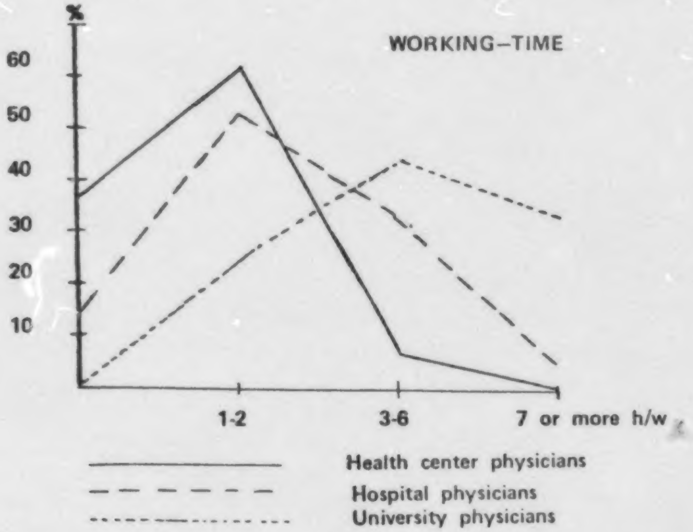


Figure 8 a

THE TIME SPENT ON FOLLOWING MEDICAL INFORMATION;
LEISURE TIME

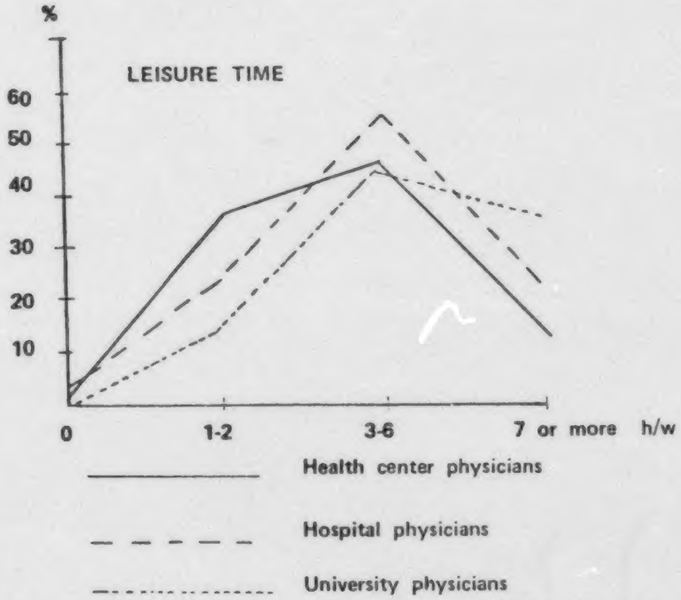


Figure 8 b

THE READING OF MEDICAL BOOKS

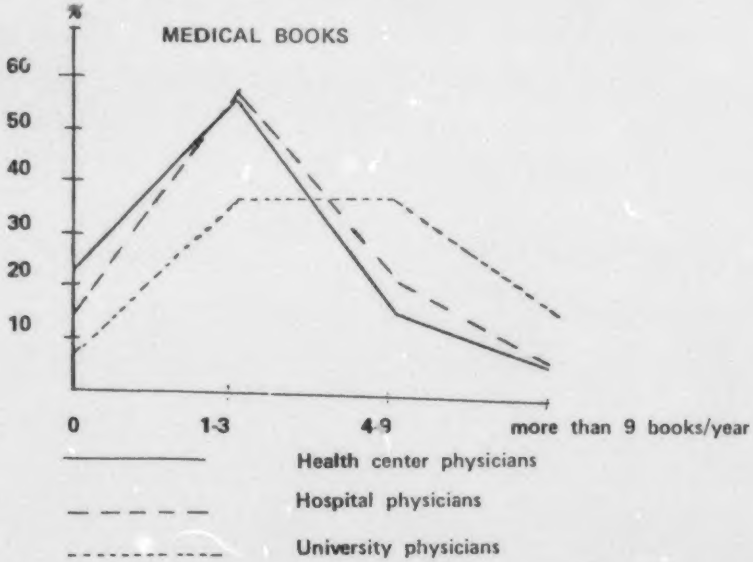


Figure 9 a

THE READING OF OTHER BOOKS

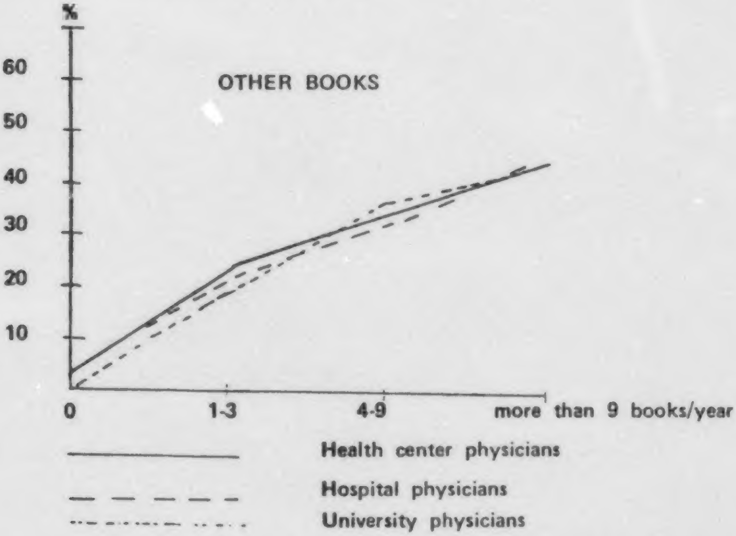


Figure 9 b

FREE-SUBSCRIPTION	**MOST FREQUENTLY READ JOURNALS**	
	Suomen Lääkärilehti	99 %
	Duodecim	97 %
	FREQUENTLY READ JOURNALS	
	Helsingin Lääkärilehti	93 %
	Nuori Lääkäri	88 %
	Medisiinari	85 %
	Lääketehtaiden julkaisut	83 %
	Akava	82 %
	Kollega	79 %
	Nordisk Medicin	79 %
	Finska Läkaresällskapets Handlingar	72 %
	World Medical Journal	71 %
SUBSCRIPTION	**SUPERFICIALLY READ JOURNALS**	45 %
	Lancet	45 %
	British Medical Journal	44 %
	Sosiaalilääketieteellinen aikakauslehti	44 %
	Läkartidningen	39 %
	J.A.M.A.	31 %
	New England Journal of Medicine	26 %
	Sairaala	23 %
	INFREQUENTLY READ JOURNALS	
	Sairaanhoitaja	15 %
	Science	12 %
	Nature	11 %
	Ugeskrift for laeger	4 %
	Tidskrift for den Norske Laegeföreningen	5 %

Figure 10

important characteristic of most regularly read journals is that they are free-subscription journals. The journals which are received by the physicians without their own initiative, or activity, are the most frequently used. Again the differences between different groups are big. Domestic, free-subscription journals are used more by health centre physicians than by the others. Only 17% of the health centre physicians subscribed to the other journals.

After realizing the importance of all kinds of Finnish medical journals, especially those which come passively without their own initiative to the table of physicians, I made a study of the structure of these journals and of their contents.

I divided the journals into three groups: professional journals, the journals of pharmaceutical industry and scientific journals.

When we examine the increase in the number of pages in Finnish medical journals from the 60s to 70s we notice that to the table of a physician come more and more pages paid by the pharmaceutical industry. The increase in the number of pages in the journals of Medical Association is 143%. And the increase in journals published by the pharmaceutical industry is 102%. The increase in the Finnish scientific journals is only 63%.

So it is true that scientific journals doubled in ten years but other journals made more than a three fold increase.

What about the contents of these three different types of medical journals?

Medical journals contain on the average 18% scientific information, 23% previously published research information and 20% advertising. In the contents analysis of the journals published by different publishers, the following unexpected findings were noted:

See Figure 11.

In scientific journals, more than a third of the total contents cover something else than research, and about

THE CONTENTS OF FINNISH MEDICAL JOURNALS IN 1975,
SCIENTIFIC JOURNALS

(N= 3406 PAGES)

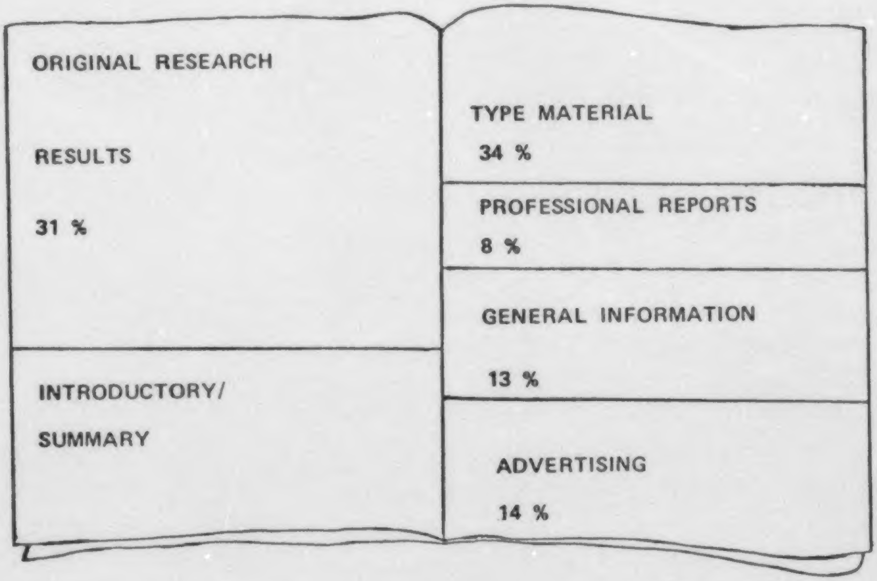


Figure 11

a sixth is advertising;

See Figure 12.

In professional journals, a fifth of the total contents is advertising and research results account for less than a tenth;

See Figure 13.

In journals published by the pharmaceutical industry, the amount of advertising is smaller than in the professional journal, and more than a third of the total contents is earlier published research information.

The publishing of original research results in domestic journals has decreased in the 1970s compared with the 1960s, and the share of introductory or summary-type material has noticeably increased. The importance of the scientific journals in publishing new research results is diminished by the fact that their size is growing more slowly than that of the other journals.

The following general conclusions can be drawn from the results of my studies. The closer the physician is to general practice, that is, the closer he is to the patient and the less a physician's work is connected with research

- the fewer university libraries and information service institutions have an impact on dissemination, the more importance other interest groups like the pharmaceutical industry have;
- the fewer scientific journals are read, the more importance free-subscription Finnish journals have in the transfer of knowledge;
- the less time there is for information reception, the less medical information is read;
- but all physicians read ordinary books as much.

THE CONTENTS OF FINNISH MEDICAL JOURNALS IN 1975,
THE JOURNALS OF THE PHARMACEUTICAL INDUSTRY
(N= 597 PAGES)

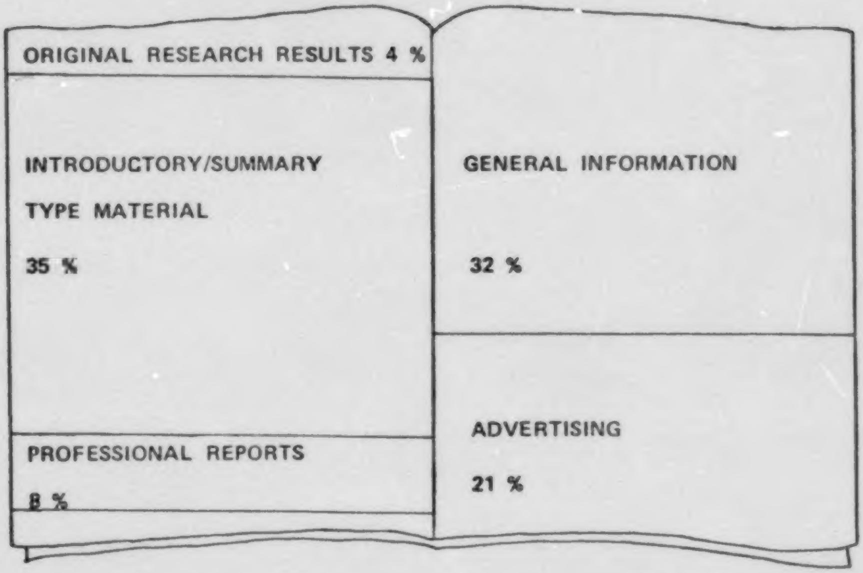


Figure 12

THE CONTENTS OF FINNISH MEDICAL JOURNALS IN 1975,
PROFESSIONAL JOURNALS
(N = 3979 PAGES)

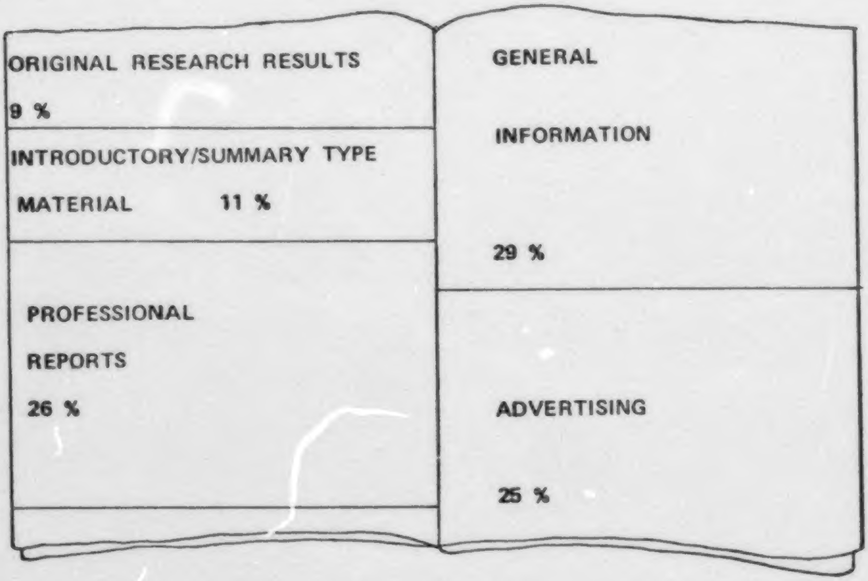


Figure 13

The generalization of the results of the contents of journals shows that the externalistic reasons for publishing are becoming more important than internalistic reasons from one decade to the other. In other words the scientific community has earlier decided more independently on the publication of journals. The increase in pages of scientific journals was smaller than the corresponding increase of professional and industrial journals. The contents of the journals also changed. The scientific content of medical journals decreased during the period studied compared with other kinds of contents of journals. Still in the 1970s advertisements, announcements, and corresponding material constituted one third of the contents in scientific journals.

The results show that the dissemination and utilization of scientific information is not only a question relating to resources or to the interaction between producers and users of information but also a question of social and economic activity in which there are competing and conflicting interest groups, and in which the social position of the physician to a large extent defines information behaviour.

I said at the beginning that I had made these empirical studies in 1978 and 1979, what has happened after that?

I am trying to develop theoretical thinking on these questions, and I have started two new empirical projects.

I hope that I will be able to report results at our next meeting.

References

1. Alestalo, Marja. Interdisciplinarity in the light of the development of science and the actual research work. International sociological association conference of the research committee on the sociology of science. Budapest, September 7-9, 1977.

2. Barnes, Barry (ed). Sociology of Science. Penguin modern sociology readings. Harmondsworth, Middlesex, 1972.

3. Ben-David, Joseph. Scientific productivity and academic organization in nineteenth-century medicine. In: Barbes, Bernard & Hirsch, Walter: The Sociology of Science. New York,

London, 1962.

4. Bernal, J.D. After twenty-five years. In: Goldsmith, Maurice & Mackay , Alan: The Science of Science. Society in the Technological Age. London, 1964, 209-228.

5. Bernal, J.D. Science in History I-IV. Harmondsworth, Middlesex, 1969.

6. Bowden, Charles L. & Bowden, Virginia M. A survey of information sources used by psychiatrists. Bull. Med. Libr. Assoc. 59 (4) Oct 1971, 603-608.

7. Brittain, J.M. What are the distinctive characteristics of information science. Paper presented at the second international research forum on information science. Copenhagen 3-6th August, 1977.

8. Choudhury, A. & Bose, A.K. Medical journalism. Journal of the Indian Medical Association, 1970, 55, 174-175.

9. Cuddihy, Robert V. Scientific information: philosophy and policies within the pharmaceutical industry. In: Day, Stacey B. Communication of Scientific Information. Basel, 1975.

10. Cummings, M.M. The responsibility of the professional society for scientific communications. American Journal of Cardiology, 1968, 21, 4-5.

11. Debons, Anthony. Foundations of information science. Paper presented at the second international research forum on information science. Copenhagen, 3-6th August, 1977.

12. Dobrov, G. Potential der Wissenschaft, Berlin, 1971.

13. Farkas, Janos. The dialectic of internal and external determinants of scientific development. In: Proceedings of the International Seminar on Science Studies. January 11-14, 1977, Espoo, ed. Paavo Löppönen. Suomen Akatemian julkaisuja 4/1977. Helsinki, 1977.

14. Forskning i kontakt med samhället. Statens offentliga utredningar 1977:55. Utbildningsdepartementet. Expertbilaga 3 till SOU 1977:52. Forskningspolitik. Stockholm, 1977.

15. Harrington, Charlene. Medical ideologies in conflict. Medical Care. November, 1975. Vol. XIII no. 11. 905-914.

16. Havelock, Ronald, in collaboration with Guskin, Alan & Frohman, Mark & Havelock, Mary & Hill, Marjorie & Huber, Janet. Planning for Innovation through Dissemination and Utilization of Knowledge. Center for Research on Utilization of Scientific Knowledge. Institute for Social Research. The University of Michigan,Ann Arbor, Michigan, 1973.

17. Hemminki, Elina. Content analysis of drug detailing by pharmaceutical representatives. Medical Education 11:210, 1977.

18. Hemminki, Elina & Pesonen, Terttu. The function of drug

company representatives. Scand. J. Soc. Med. 5: 195-114, 1977a.

19. Hoel, Ivar A. On the concept of knowledge and its relation to information science. Paper presented at the second international research forum on information science. Copenhagen, 3-6th August, 1977.

20. Houghton, Bernard. Scientific Periodicals. Their Historical Development, Characteristics and Control. London, 1975.

21. Howard-Jones, Norman. Our medical literature - then and now. British Journal of Medical Education. Vol. 7:70-85, 1973.

22. International Sociological Association Conference of the Research Committee on the Sociology of Science. Budapest, September 7-9, 1977.

23. Jensen, Anders Oltar & Jensen, Hans Siggaard. Medicinsk Videnskabsteori. København, 1976.

24. Kaczmarek, Jan. On Principles of Science Policy. Problems of the Science of Science. Ossolineum, 1971. Second special issue of the Polish quarterly Zagadnienia Naukoznawstwa.

25. Klaukka, Timo. Lääkärien saama lääkeinformaatio. Suomen Lääkärilehti, 1975:30.

26. Kotre, John N. & Mann, Floyd C. & Morris, William C. & Vanselow, Neal A. The Michigan physician's use and evaluation of his medical journal. Michigan Medicine, 1971, 70, 1. 11-16.

27. Krieger, Abbott J. & Lyle, Robert S. Information sources for physicians. Postgraduate Medicine. Vol. 56 No. 4, October, 1974. 143-146.

28. Kunz, Werner (ed). Informationswissenschaft, Stand, Entwicklung, Perspektiven - Förderung im IUD - Programm der Bundesregierung. Ergänzer und erweiterter Bericht der 1. Fachtagung Heidelberg, Juni 1976. München, Wien, 1978.

29. Kunz, Werner & Rittel, Horst W. & Schwuchow, Werner. Methods of Analysis and Evaluation of Information Needs. München, 1977.

30. Löppönen, Paavo. Tieteen tutkimus-muistio. Suomen Akatemia. Helsinki, 1975b

31. Meadows, A.J. Communication of Science. London, 1974.

32. Michaijlov, A.J. & Černyj, A.I. & Giljarevski, R.S. Informatik 1 ja 2. Grundlagen, DDR, 1967.

33. Navarro, Vicente. The fetishism of industrialization. A critique of Ivan Illich. Monthly Review. Vol. 28 (1976) No.5. 34-46.

34. Neveling, Ulrich & Wersig, Gernot. The information scientist

of the 1980's Europe. Paper presented at the second international research forum on information science. Copenhagen, 3-6th August, 1977.

35. Novikov, E.A. & Jegorov, U.S. Informaatio ja tutkija. Neuvostoliittoinstituutin julkaisusarja B 21. Helsinki, 1976.

36. Nowakowska, Maria. Microparadigms - simulations - prognoses, further remarks on the theory of epidemical spread of scientific objects. Problems of the science of science, 1972-1973.

37. Ostrowska, Teresa. Wartość informacyjna czasopism medycznych w świetle wypowiedzi użytkowników zagadenienia informacji naukowej pan 21 (1972) no 2.

38. Pfetsch, Frank & Zloczower, Avrahem. Innovation und Widerstände in der Wissenschaft. Wissenschaftstheorie. Wissenschaftspolitik. Wissenschaftsplanung. Band 19. Düsseldorf, 1973.

39. Price, Derek de Solla. *Science since Babylon*. Enlarged edition. New Haven and London 1975.

40. Rehu, Marjatta. Lääketehtaiden lääkäreihin kohdistamasta lääkeinformaatiosta. Liiketaloustiede II:n laudaturtutkielma. Kevät, 1973. Turun kauppakorkeakoulu.

41. Roland, Charles G. & Kirkpatrick, Richard. Time lapse between hypothesis and publications in the medical sciences. *The New England Journal of Medicine*. June 12, 1975, 1273-1276.

42. Savolainen, Reijo. Kirjastotiede tieteenä. Tampere, 1978.

43. Schor, Stanley & Karten, Irving. Statistical evaluation of medical journal manuscripts. *JAMA*. March 28 1966. Vol. 195: No. 13:145-150.

44. Searle, Shayler R. The publish - or - perish syndrome. EEE transactions on professional communication. September, 1973.

45. Vickery, B.C. An approach to information science. Paper presented at the second international research forum on information science. Copenhagen, 3-6th August, 1977.

46. Wilson, Thomas & Stephenson, James. Dissemination of Information. London, 1969.

47. Ziman, John. *The Force of Knowledge. The Scientific Dimension of Society*. Cambridge, 1976.

PROJECT ATHENA, "THINKING ALOUD", AND USER-ORIENTED RESEARCH

Kathleen T. Bivins

1. Introduction

Historically, in terms of IRFIS conferences, the combined issues of methodology and of user-foci have been significant topics. This particular conference, IRFIS 4, has as co-themes 'User-orientated research in information science' and 'Experimental and practical investigations of the mediation and use of information'. These are obviously topics of continuing concern and interest, not only for IRFIS, but also for national funding bodies in the information science area (e.g. the National Science Foundation of the United States, the Delegation for Scientific and Technical Information of Sweden, and the British Library Research and Development Department of Great Britain). It should be remarked here that one of the troubling problems for information science researchers in the particular area of user research (as well as in other areas) is the well-known basic-applied continuum. It is interesting to note that a number of the IRFIS participants have swayed from more theoretical to more applied work over the years since the first IRFIS meetings. It is unclear as to whether this possible trend makes a statement about research direction or is more representative of a group of researchers trying different approaches. Whichever point of view is more nearly the case, it is certainly true that the IRFIS conferences have influenced all those participating.

2 Goals and definitions

2.1 Definition of user-oriented research

One can be ridiculously broad in defining user-oriented research.

It is rather obvious that in the field of information science, everything is ultimately concerned with users, and that all research is therefore 'user-oriented'. Needless to say, for purposes of this paper, the definition of such research has been rather severely limited. User-oriented research, in this context then, consists of the range of activities undertaken in direct searches for information. While there are many more specific activities and charts, diagrams, and the like, of these activities, here the concern is with the following:

- information needs surveys/studies
- the search process
- catalog use studies

2.2 What do users 'think'?

The current information retrieval systems, whether card catalogs, printed book indexes, online bibliographic databases or whatever, are by far and large based on a series of assumptions. As is well-known, we do not know how users 'think' and, therefore, we do not know either the best means of designing systems, or of using current ones, so as to provide the most useful means of access to stored information. The methods now in use, such as boolean searching of digitally stored files of references, author-title access to catalogs, thesauri and subject-heading lists of various types, and their particular arrangements, are all but a series of guesses or assumptions as to how things should be organized. Some of these assumptions may be well-founded, some may not. In any event, we need more information ourselves on how to provide information.

2.3 A question

Following from the above, an obvious question is raised. Are there any valid means of attempting to discern 'thoughts' and/or modes of behaviours of users?

2.4 An answer

The answer is yes - in a limited sense. The following list contains the most commonly used methods/techniques of attempting to answer our question:

- observation
- inquiry
- guessing (very common!)
- 'thinking aloud'

2.4.1 Examples of techniques

> "I checked it very thoroughly", said the computer, "and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never actually known what the question is".
>
> Douglas Adams
>
> "The Hitchhiker's Guide to the Galaxy"

This quote illustrates quite well the status of our efforts in this field. While this paper makes absolutely no attempt to be comprehensive, nor does it pretend to be a state-of-the-art or review article in any sense, it does present a series of observations and opinions about what we are doing now (in attempting to answer our question) and what is a promising way to proceed.

Let us return briefly to the listed techniques. Observation is an interesting technique in that one can acquire a fair amount of data of the 'counting' sort. For example, one can observe the activities at the reference/information desk of the library and note that 'n' people approached the desk, over a period of 'n' hours, taking an average of 'n' minutes/seconds each. Furthermore, there were 'n' women, 'n' men, 'n' children, and the reference librarian had to be prompted by an 'ahem' or other such noise half the time (or not at all, or always!). These kind of data certainly tell us something - whether they are an aid in system design or not is more questionable.

The inquiry technique makes use of the very well-known questionnaire/survey approach. Of course, frequently methods are combined, but these are just simple illustrations of the techniques. Questionnaires again gather all sorts of data and, generally, are analyzed after the fact for presentation of the data in statistical form. To return to our reference desk, each person approaching is asked to fill in a questionnaire, perhaps of the 'it'll only take a minute to fill out' type. The form asks for data on, e.g., age, sex, education, use of library(ies), why is individual using reference desk, has the card catalog been tried, and so on. Statistical analysis of the information on 'n' questionnaires shows that 'n' percent were 'n' sex, and so on.

Guessing. This scarcely seems a method or a technique, but it is a tried and true form, and is in wide use. A more formal name might be the trial-and-error method, although trial-and-error implies trials, detection of errors, correction of same, and retrials. Users may indeed (and often do) use this approach, but those who have labored to devise systems are often impelled to guess as to how best to proceed, since so little is known about why, what, and wherefore. Which brings us to....

2.4.2 'Thinking aloud'

For the uninitiated, thinking aloud requires some explanation. Thinking aloud is not necessarily equated with muttering to oneself on the street, sudden exclamations, sotto voce comments during lectures, and the like. It does however, bear a relationship to these and in fact, is derived from research by a Soviet psychologist, Lev Vygotsky, into the verbalizations of children[1]. Vygotsky was interested in the process of externalization-internalization of cognition. He observed that young children 'thought aloud', but that 'thinking' gradually became internalized as the child matured.

Over the past 15-20 years, there has been sporadic and then increasing interest in using this method (or developing variations thereof) for attempting to find out what people 'think' under various circumstances. That is, if one can induce persons to verbalize, under presumably controlled conditions, then perhaps one has data which couldn't otherwise be collected. There are obvious and glaring problems - the rather extreme obtrusiveness, the possibility that verbalization per se changes the 'thought', and so on. However, in psychological circles it is considered a valid method, and most interestingly for our purposes, it has been used, quite successfully, a number of times in our field.

Studies in this area consist of a combination of theoretical and experimental work, aimed, more or less, at an understanding of what happens during the search process. A number of names can be cited here. Carlson[2] was probably the first in this field to use the technique. Much later can be noted the work of Ingwersen, Timmerman and Kaae[3,4], Mark Pejtersen[5], Keen[6], Belkin and Oddy[7], and Bivins[8,9]. The general overall method is to use a tape recorder to record 'thoughts' and then to analyze the data so collected.

The work of Ingwersen et al. is certainly the most thoroughly performed and documented in this area, and, in my opinion, should serve as a model and a pattern for future and current researchers. Their research, at the Danish Library School, was primarily concerned with finding out something of patterns in search methods/routines of public librarians and/or students searching in public libraries, with an eye towards improved educational methods in the library school. Their careful work and extensive data analysis has led to more experiments and has shown promising leads for future studies.

Mark Pejtersen, also of the Danish Library School, using the same methodological technique, was able to develop a new kind of classification scheme for fiction in public libraries. Based on analyses of verbal protocols (obtained through tape recordings), she developed a scheme which involved types and characteristics of fiction works, and which was used by librarians in directing users to what they wanted. An unexpected side result of this was that many of the books which had **not** previously circulated were borrowed by readers! This is a small, but convincing, piece of evidence for developing user-oriented systems. Readers really borrowed books which were classed (in a multi-dimensional way) through a scheme developed from their (the readers') wants and needs.

Belkin and Oddy have been working on a design study for a system which would be based on a model other than the traditional one. They first postulate that information retrieval system users have an anomalous state of knowledge (ASK), whereby they cannot specify exactly what they need, and therefore can't be expected to 'match' that ASK to the specific terms required by an indexing language. Once again, using the thinking aloud technique, users of a particular library were queried as to their reasons for coming there. The recorded information was transcribed into verbal protocols, which were then subject to rather rudimentary linguistic analyses, resulting in the development of what can be called association networks revolving around the central concepts expressed in the protocols. Next, the subjects queried were asked to verify the networks, which were accordingly adjusted. In order to attempt to construct a novel retrieval system, Belkin and Oddy performed a similar linguistic analysis on abstracts of certain articles , sent the resulting networks to the authors, and again, made adjustments as necessary. They then had the beginnings of what would be an information retrieval system - document concepts expressed as networks of associations, and representations of verbal protocols expressed in a similar way. It is obvious, of course, that the proposed system would attempt to develop an ASK of the user (through word co-occurrence strenghts and so forth) and then search its other networks for, presumably, relevant information. It will be most interesting to see whether, in fact, and how, if so, such a system would actually work.

3 Project Athena

Project Athena was an investigation undertaken at the University of California, Los Angeles, Graduate School of Library and Information Science (UCLA, GSLIS) by Bivins, with the assistance of research students. The background for this work was fuelled by a combination of factors: general frustration with subject indexes, curiosity about the thinking aloud method, and an awareness of the work of cognitive psychologists, Rosch[10] in particular, on the process of 'basic level' categorization. By this is meant that within structural categories (e.g., furniture, chair, armchair), a more or less mid-point exists, a basic level, at which one finds maximum within-category similarity, and minimum between-category similarity. In the example of furniture, chair, armchair, chair is the basic level object. This type of categorization is argued by many researchers (Rosch et al.) as being a fundamental human process.

3.1 Objectives

Having long had an interest in indexing, information retrieval systems, and the like, it was a natural progression to wonder whether, in fact, users of index lists had 'basic level' categories in mind and whether, also in fact, the given index language aided or hindered that categorization. Thus, the objectives of Athena can be stated as follows:

In a subject search investigate:

1. Level of indexing expected by users versus actual level used in source.
2. Cognitive processes occurring during
    - formulation of search strategy
    - actual search

Next, how does one begin to tackle this? A decision was made to perform an initial using first-year library school students, none of whom had been exposed to courses in subject cataloging, classification, and the like, and all of whom could be expected to participate as part of a class requirement, thus eliminating possible 'volunteer' problems. It was further decided to concentrate primarily on the first object (level of indexing) and if anything of the second objective emerged during the study, it would be a bonus. Thus, we can define the subjects:

For pilot study:
First-year library school students

For future study:
- reference librarians
- naive users

The obvious problem of using library school students as opposed to any other group was thought to be somewhat countered by the facts that this was to be a pilot study, methodological problems were of strong interest, and that the students were a well-defined group.

3.2 Method

The following table outlines the method used to attempt to determine something of the level of indexing expected by users versus that used in the source:

1. Each student gets same query.
2. Student develops list of search terms - thinking aloud into tape recorder.
3. Student performs search at card catalog - thinking aloud into tape recorder.

The query, which was developed after considerable testing and discussion of queries, met the needs of the study in that it could be answered in the source library utilized, it satisfied requirements for the existence of a range of subject terms, and it was relatively clear and simple.

> Query:
> I am interested in setting up a hotline as part of a suicide prevention program. What books/documents does the library have about such hotlines and the experience of their operators?

At this point, a little background on the library system in use is helpful. First, the study took place at UCLA, which contains a large research library. This library uses a divided card catalog as its major access to the collection. The card catalog has separate sections for author-title entrypoint, and for subject entrypoint. The subject index system used is the Library of Congress Subject Heading list.

As noted in the preceding remarks, the students (34 in number) were presented with the query, asked to develop search terms that they thought would be useful, then subsequently presented with that list (as derived from their narrative) and asked to start searching for documents answering the query in the UCLA Research Library card catalog.

Before the first session with the tape recorder, the students had been told generally about the study, and specifically about 'thinking aloud' and tape recording. They had not been informed about the level of indexing hypothesis. The initial tape recording session caused few problems, and, most notably, there was a lack of self-consciousness among the students. Each student was given an appointment of 10 minutes in length, and met with a research assistant in order to read the query, and then to start verbalizing in response to it. The following are some excerpts from transcriptions of the interviews:

> "Right now? OK, I'm ___________. The subject headings that I would look under for this particular topic would probably be something along the lines of suicide aid or medical prevention or as to whether there is a specific way of setting up a hotline like possibly setting up help programs. Something under sociology, maybe something like that. Maybe another avenue would be to look under psychology, because suicide is part of psychology... Maybe under toxicology, because there are toxicology hotlines already set up, as I said, sociology..."
>
> "OK, my name is ___________. Think about suicide prevention programs, personnel of such programs. General headings... could be psychology, abnor-, not abnormal psychology, but probably psychology... Perhaps something like outreach in the sensen of reaching out to people who are in need..."
>
> "My name is ___________. I would probably look first under suicide prevention and if I didn't find anything there, or I didn't find what I was looking for there, I would look under community services. I wouldn't expect to find anything under hotline, that being too narrow a topic, and I would probably want to see how I could get at community service angle by looking at the LC subject headings..."

Transcription of the tapes (a laborious, frequently difficult process) resulted in the production of lists of terms for each student as expressed in the narrative. The underlined words in the examples of narratives illustrate this. The students then met with a research assistant at the card catalog of the Research Library, and were then given their list of terms. Once again they read the query, and were asked to search in the card catalog for references to documents they thought would answer the query. They were to use their list of terms as a start point, but could continue as they wished, and/or as led by the card catalog. The subject was further asked to verbalize, insofar as possible, his/her thoughts and activities while doing this task. The following are excerpts from transcriptions of these sessions:

> "OK, this is _________ and my primary term is suicide, so I'm going to look under that first... Wow, there's a lot of cards under suicide... I'll start at the beginning...suicidal behavior, hmmm, just the term suicide by itself, I don't know if that's going to refer mer to anything on hotlines. Hmmm, maybe something after suicide... suicide prevention, that sounds... might be something under that... still nothing directly about... nothing about hotlines, crisis centers, (while looking under suicide prevention)..."

> "I'm back, and I'm going to start with suicides, or suicide singular, and then move on from there. So as we head toward the 's' file in the catalog, which is far enough away to not be useful and I suddenly realize I guess the size of these things is incredible...How do you spell suicide? Suicide in Russian, OK... Sure is a lot in here about survivors of suicide, it's interesting, but not much to do with this topic. You wonder about the cataloging of these books, if they really have what you want in them but you don't know..."

3.3 Analysis

The two sessions of tape-recording produced voluminous amounts of typed transcriptions. Analysis of these is difficult, but in this case, indexing terms (proposed and actual) were the focus of interest, and could be extracted without too much difficulty. As noted before, students had a list of their original proposed search terms, and subsequent lists of their actual used terms (in the card catalog) were developed from the second taping session. Students at the first session almost invariably chose 'suicide' or 'suicide prevention' as primary search terms. In case of doubt, they generally decided to search at a broader level, such as under 'psychology', but rarely even thought about a narrower level, despite the fact that the Library of Congress Subject Heading list instructs indexers to use the most specific term possible. In the case of the query, that term would be 'hotline'. This is an acceptable index term in the list, so its use or possible non-use cannot be ruled out on the basis of non-existence as an index term. The students were not technically familiar with the LC Subject Heading list, but all were users of the Research Library and had used other U.S. research libraries, the vast majority of which use LC Subject Headings. In fact, although there were a number of documents in the Research Library which may have satisfied the query more or less, there was one particular book which answered the query in every respect, and which was indexed under one term only - 'hotlines'. It was not located by any of the students. One can say that the analysis of the tapes yielded the following results:

1. Near unanimity of a choice of primary search term:
   'suicide'
   or
   'suicide prevention'
2. Rejection of direct and specific term:
   'hotlines'

This confirmed, within the limitations of the study, that a basic, or mid-range, level of indexing is used by subjects (the students, in this case). However, this level of indexing was not that which was used in the indexing itself. It must be kept in mind that one cannot generalize at this stage, tempting as it may be, but that the results of this study certainly lead one to believe that this is an area to investigate with a larger population group, a more varied one, and with multiple queries, and differing index languages. General comments on the study can be listed as follows:

1. Minimal self-consciousness apparently caused by 'thinking aloud' procedure.
2. Procedural controls need finer definition.

3. Broader experience range of population groups as subjects is needed.
4. Appropriate analytical tools are needed.
5. Results are promising for future development of query-based/user-responsive indexing.

In regard to analysis, a number of possibilities for the future present themselves:

1. Map relationships between terms proposed by individuals.
2. Compare term usage and relationships between individuals.
3. Map relationships of terms during search process.
4. Identify implicit 'success' criteria in search/decision-making process.

3.4 Conclusions

In sum, in regard to this study, it is possible to state that the initial results are of interest and that further testing and refining may well give more concrete evidence supporting the level-of-indexing hypothesis.

4 General conclusions

Having discussed at some length the 'thinking aloud' technique, and one study in particular, it is now time to consider whether in fact this method should continue to be used. Thinking aloud is fun, and somewhat novel, but more to the point, it seems to be yielding bits and pieces of information which have not previously emerged, and/or which support bits and pieces from other types of studies. One of the possible problems relating to thinking aloud is connected more with the tendency to try to do too much within one study, than with the method itself. A more serious problem is the actual data analysis of verbal protocols. This certainly needs a great deal more work, and perhaps some rather innovative thinking. Procedural problems in use of tape recorders, self-consciousness, and that sort of thing, are less difficult to resolve. On balance, these are somewhat minor quibbles.

The method itself is extremely promising and certainly deserves continued use in this field.

References

1. Vygotsky, L.S. Thought and Language. Cambridge, MA, MIT Press, 1962.

2. Carlson, G. Search Strategy by Reference Librarians. Sherman Oaks, CA, Advanced Systems Division, Huges Dynamics, 1964.

3. Ingwersen, P., Johansen, T., and Timmerman, P. User-librarian negotiation and search procedures: a progress report. In: Harbo, O. and Kajberg, L., eds. Theory and Applications of Information Research. London, Mansell, 1980, p. 160-171.

④ Ingwersen, P. and Kaae, S. User-Librarian Negotiations and Information Search Procedures in Public Libraries: An Analysis of Verbal Protocols. Final research report. Copenhagen, School of Librarianship, 1980.

5. Mark Pejtersen, A. Design of a classification scheme for fiction. In: Harbo, O. and Kajberg, L., eds. Theory and Applications of Information Research. London, Mansell, 1980, p. 146-159.

6. Keen, E. Michael. On the processing of printed subject index entries during searching. Journal of Documentation: 33:4 (1977), p. 266-276.

7. Belkin, N.J. and Oddy, R.N. Design Study for an Anomalous State of Knowledge Based Information Retrieval System. Final report on Grant S/1 SG 09, 2May78-15June79, to the British Library Research and Development Department, September 1979.

8. Bivins, K.T. Users of information retrieval systems: a frame-theoretical approach. Proc. 3d International Research Forum in Information Science. Oslo, Statens Bibliotekskole, 1979.

9. Bivins, K.T. User/systems interaction analysis. International Forum in Information and Documentation (forthcoming).

10. Rosch, E. Universals and cultural specifics in human categorization. In Brislin, R., Bochner, S. and Lonner, W., eds. Cross-Cultural Perspectives on Learning. New York, Halsted Press, 1975.

THE CONCEPT OF ENTAILMENT FOR THE PURPOSE OF MEDIATION

Marta Grabowska

Descriptions of the content of documents, made for the purpose of scientific information can be classified into two groups:

1. those written in a natural language,

2. those written in an artificial language, i.e. an information retrieval language.

Group 1 includes mostly summaries (written as well as oral) while group 2 - UDC symbols, descriptions based on key words and uniterms, and all descriptions making use of descriptors (as parts of a standardized vocabulary). In this paper I shall be solely concerned with descriptions in a natural language.

Standards in the various countries[1], as well as recommendations of international organizations[2], list several kinds of description of the content of documents in a natural language. These are mostly written summaries, but their definitions as well as their classes are in most cases not disjoint or based on different criteria. I shall endeavour here to suggest a definition of the concept of description of the content of document in a natural language on the basis of the concept of entailment.

Formalism and definitions

A text X may be treated either as a sequence of sentences (sentence equivalents being here, for simplicity, taken on a par with sentences) or as a set of sentences, i.e. a number of sentences without repetitions[3].

It can also be treated as a sequence of words. (The concept of text as a set of words, i.e. a number of words without repetitions, which is also possible, will be disregarded here). The text X treated as a set of sentences will be denoted by X'; treated as a sequence of sentences by D/X/; and treated as a sequence of words, as d/X/. D and d are functions which assign to every text a certain natural number. Note that $\bar{X}' \leqslant D/X/$ and $D/X/ \leqslant d/X/$. $\bar{\bar{X}}' = D/X/$ only if no sentence occurs in X two or more times; $D/X/ = d/X/$ is an improbable case of a text consisting of one-word sentences only (this is grammatically possible in Polish, but most unlikely in English, except perhaps for short poems consisting of sentence equivalents, the case may be disregarded when it comes to documents in library and information science). Note also that for two different texts, X_1 and X_2, it may occur that $D/X_1/ < D/X_2/$ while $d/X_1/ > d/X_2/$ if, on the average, X_1 consists of longer sentences then X_2 does.

Further, E/X'/ will be used to denote the set of the sentences entailed by X'. The original, Tarskian, concept of consequences[4] is not suitable for our purpose, because in Tarski's system every sentence implies a transfinite number of sentences, and hence has a transfinite number of consequences, in view of the fact that $p \Rightarrow p \vee q$, $p \Rightarrow p \vee r$, etc., and also $p \Rightarrow /q \vee \neg q/$, and $p \Rightarrow /r \vee \neg r/$, etc.. The concept of entailment to be used here might be that of Parry's analytic implication[5], or the entailment system formulated by Zinovev[6], or Epstein's[7] relatedness logic. The final choice need not be made for the time being, before the advantages of the various systems of entailment have been compared from the point of view of our needs.

X was at first made to stand for any text, but we shall now substitute for it the symbol O = original, C = extract, P = paraphrase, T = translation. To indicate that we mean, in a given case, a certain original and its extract, not just an extract of any text, we shall add subscripts linking the symbols in pairs, e.g. O_i and C_i means that C_i is an extract of the original O_i. Should we consider several extracts of one and the same original, we would develop the subscript of C into C_{i_1}, C_{i_2}, etc.

To indicate that a text is in a given language we shall use the sentence-forming function L with a subscript; thus $L_j/X/$ says that X is a text in a language j.

Now, for a text to be an extract of another text it must meet the following conditions:

1. $d/C_i/ \ll d/O_i/$,

2. $E/C'_i/ \subseteq E/O'_i/$,

3. $\bar{\bar{C}}'_i = D/C_i/$,

4. there is a language j such that $L_j/O_i/$ and $L_j/C_i/$.

Now, 1 states that an extract must be much shorter than the original, 2 says that the set of sentences entailed by an extract must be included in the set of the sentences entailed by the original, 3 indicates that an extract should be a sequence without repetitions (this is a practical rather than a theoretical requirement), and 4 says that the two texts must be in the same language.

These conditions can be linked by conjunction signs to form the definiens of a formal definition, as follow:

$$C = (d/C_i/ \ll d/O_i/) \wedge (E/C'_i/ \subseteq E/O'_i/) \wedge$$
$$\wedge (\bar{\bar{C}}'_i = D/C_i/) \wedge (L_j/O_i/ , L_j/C_i/)$$

but for our purposes it is more convenient to treat them separately (the same procedure will be observed in the case of other definitions).

If a text is to be a paraphrase of another text it must meet the conditions listed below:

5. $d/P_i/ \simeq d/O_i/$,

6. $E/P'_i/ = E/O'_i/$,

7. $D/P_i/ \simeq D/O_i/$,

8. there is a language j such that $L_j/O_i/$ and $L_j/P_i/$.

This is to say that a paraphrase must be roughly of the same length as the original (and possibly, which is a debatable point and is indicated by condition 7, must be divided into roughly the same number of sentences, which is more a stylistic requirement), must entail the same set of sentences as the original does, and must be in the same language as the original.

To be a translation of another text a given text must satisfy the following conditions:

9. $d/T_i/ \simeq d/O_i/$,

10. $E/T_i'/ \simeq E/O_i'/$,

11. there are languages j and k such that $j \neq k$ and $L_j/O_i/$ and $L_k/T_i/$.

Thus, the two texts must be of approximately the same length and must entail roughly the same sets of sentences, the last-named condition is formulated less rigorously than in the case of the paraphrase because usually some meaning of the original is lost in the process of translation, or on the contrary, the translator - mainly for linguistic reasons - has to interpret the original and may thereby amplify its meaning. Finally, the two texts must be in different languages.

Categorization

For the purpose of the typology of characteristics of the content of documents in a natural language we note that they may contain sentences of the following types:

1. sentences which are drawn verbatim from the original, i.e. in a sense extracted from it (to be denoted by ex),

2. sentences which are paraphrases of sentences to be found in

the original (to be denoted by par_{ex}),

3. sentences which are translations of sentences extracted from the original (to be denoted by tr/ex/),

4. sentences which are translations of paraphrases of sentences occuring in the original (to be denoted by tr/par_{ex}),

5. sentences which entail from the original according to one of the entailment systems mentioned above (to be denoted by en),

6. sentences which are translations of sentences entailed by the original (to be denoted by tr/en/),

7. sentences which are paraphrases of sentences entailed by the original (to be denoted by par_{en}),

8. sentences which are translations of paraphrases of sentences entailed by the original (to be denoted by tr/par_{ex}),

9. moreover, a given characteristic may contain additional information, e.g. sentences which contain additional information about the content and/or form of the original, i.e. _associated sentences_ denoted by com (com_c = about the content, com_f = about the form of a document); strictly speaking, the _com_ sentences do _not_ form part of a characteristic of the content of a document in the sense defined above because they are a kind of commentary and consist of statements which neither occur in the original nor entail by the original.

It is self-evident that only _ex_ sentences are drawn from the original(s). Note also that _com_ sentences should never form a complete characteristic themselves, but should always be accompanied by sentences belonging to at least one of the remaining eight categories. The _com_ sentences are metatextual in character (in the sense of forming a text that refers to another text) and carry bibliographical information and possibly also information that comes to a (brief) review.

It can also be seen easily that not all the combinations of the categories listed under 1-9 above have their special terms in the literature of the subject. In many cases one such term covers several such combinations, whereas in other cases such combinations are not given any particular names at all. Thus, for instance, the term *extract* usually denotes a characteristic consisting of sentences drawn from the original (whether in the same order as they occur in that original or not), while characteristics consisting of paraphrases or translations of sentences drawn from the original, and those which consist of some extracted, some paraphrased and some entailed sentences - also possibly with *com* sentences, are *not* assigned any specific names. Yet it seems that it is imperative to base the classification of characteristics of documents in a natural language on the type of the sentences of which they consist, because this criterion indicates well the relation which a given characteristic bears to its original. When making a correct use of characteristics in a natural language we have to realize the role played by the operations of extracting, paraphrasing, translating, entailing and commenting, performed on the original. It is to be emphasized in this connection that the operations mentioned above performed by a human being can easily blur the differences between them because all of them are roughly treated as 'mental'. It is otherwise when it comes to their automation: extracting on the basis of specific criteria already gives good results, while paraphrasing and translating are far from being satisfactory: entailing according to various systems of entailment is still examined (also in Poland at present), but commenting seems to be impossible for the time being. Combining extracted with paraphrased and entailed sentences into a coherent whole is still also beyond the scope of automation.

We thus have the following possibilities in the structure of characteristics: $ex\ \underline{v}\ en\ \underline{v}\ par_{ex}\ \underline{v}\ (ex\ \underline{v}\ en)$... etc., i.e.

$$\sum_{k_1\ldots 8}^{n\,=\,8} = \frac{n(n-1)\ldots\,(n-k+1)}{k!} = 255 \qquad \text{or } 2^8 - 1 = 255$$

since in each case the characteristic can be accompanied by $com_c\ \underline{v}\ com_f\ \underline{v}\ com_{(f+c)}$. This shows the great variety of characteristics which are not only possible but also encountered in practice, although this variety of types is not reflected in the current terminology.

There is finally the issue of the semantic value of the characteristic. We usually try to inform one about the main theses formulated by the author of the original or convey essential information about its subject matter. This condition is not met by the definition above which merely requires that what is entailed by the characteristic should be included in what is entailed by the original. This definition would accordingly be satisfied by a characteristic that raises minor or quite inessential points mentioned in the original. Self-evidently, this semantic requirement does not lend itself easily to formalization, and no endeavour will be made here in that direction.

However, it is to be noted in this connection that Witold Marciszewski in his Metody analizy tekstu naukowego (Methods of Analysis of Scientific Texts)[9] suggests that a characteristic is to be considered satisfactory if each of its sentences could be used as a title of a section of the original. His geometrical interpretation, which will not be discussed here, requires that the order of the sentences in the characteristic should be the same as the order of the corresponding sections (or possibly paragraphs) of the original. He refers to this as a "picture of the original". W. Marciszewski also introduces the concept of "a selective excerpt of the original"; by which he means a set of sentences that are suitable for titles of some sections of the original only, without covering the whole of the original text. The characteristic proper (in his sense) should, it seems, be used both as the characteristic which accompanies the original text and as that which is to be included in information bulletins for addressees who are not strictly defined. Selective excerpt could be used rather as sources of information for well-defined groups of addressees.

At the present state of semantic studies any further research for more rigorous formulations seems futile.

References

1. for example: USA ANSI Z.39.14-1971 Writing abstracts; USSR

GOST 7.9-70 Sistema informacjonno-bibliografičeskoj dokumentacji. Referat i annotacija; Poland PN-77/N-01221 Adnotacje i analizy dokumentacyjne.

2. ISO 214-1976 Documentation. Abstracts for publications and documentation; SEV RS 2132-99 Pravil'a podgotovki rezjume; SEV RS 2133-69 Pravil'a podgotovki referatov; UNESCO-SO/NB-5 Guide for Preparation of Scientific Papers for Publication.

3. The stipulation "without repetitions" means that there are no equiform (or synonymous) elements in X'. It can be more rigorously stated thus:

   For any text X which is a string of sentences $z_1 \ldots\ldots z_n$

   (i) for any x in X' there is an $1 \leqslant i \leqslant n$ such that

   $E(z_i, x)$, where E stands for the relation of equiformity,

   (ii) for any x, y in X', if $x \neq y$ then $\neg E(x,y)$.

   The consequent in (ii) can be replaced by $\neg S(x,y)$, where S stands for the relation of synonymity. The latter is a stronger condition than the former, but it seems debatable whether it is necessary.

4. A. Tarski: Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften. Monatshefte für Mathematik und Physik, Vol. 37,1930; A. Tarski: Grundzüge des Systemenkalküls. Fundamenta Mathematicae Vol. 25, 1935 pp. 502-526; A. Tarski: Über einige fundamentale Begriffe der Metamathematik. Sprawozdania Towarzystwa Naukowego Warszawskie III, Vol. 23, 1930.

5. W. Parry: Ein Axiomensystem für eine neue Art von Implikation (Analytische Implikation). Ergebnisse eines mathematischen Kolloquiums. Vol. 4, 1933.

6. A. Sinowjew, H. Wessel: Logische Sprach Regeln. Eine Einführung in die Logik. Berlin 1975.

7. R.L. Epstein: Relatedness and implication. *Philosophical Studies*. Vol. 36 no 2, August 1979, pp. 137-173.

8. for example: L.Ch. Bernier: Abstracts and abstracting. In: *Encyclopedia of Library and Information Science*. Ed. A. Kent [and others]. Vol. 1. New York, 1968, pp. 16-38.

9. W. Marciszewski: Metody analizy tekstu naukowego. Warszawa, 1977.

LULID5/LIBR-1065

FACTUAL DATA RETRIEVAL FROM A DATA BASE CONSTRUCTED ACCORDING TO THE SAP-TECHNIQUE. - A PROJECT INCORPORATING GOVERNMENT COMMITTEE REPORTS

Björn V. Tell & Irene Wormell

> -Intelligence is defined as the organised capability of a social system - a firm, organisation or a nation - to deal with its internal and foreign problems by procuring and using open and secret knowledge.
>
> Steven Dedijer

Government information for legislative action

Without accurate information about the current states of problem areas, general guidance and control of a society are not possible. National policy and new legislation are prepared by thourough studies made by government committees which try to explore the problems from broad viewpoints. In accordance with political writers, such as Dror (69,74), Harden (80) et al., we believe that the operation of the political decision-making organs, the Government and the Parliament, could be improved by a more efficient information input.

In the science fields a comprehensive and well structured knowledge base has been developed. Around 500 publicly available data bases are at the disposal through electronic networks for those who have access to terminals. One reason for the ever-increasing efficiency of science and technology is that, whenever a new venture is undertaken, the results of past efforts are systematically retrieved and utilized. In the political area, the dissemination system of gathered data is poor, compared with the science fields. In the policy-making area the Government sets up committees to review past efforts, but the systematization lacks and earlier results are often not utilized because earlier data collections are not readily accessible to the investigators. There is a failure to develop knowledge data bases by using the modern technology. Government research reports are just published and put on the market without further ado. The recommendations of the committees are often reviewed and put into a bill to the Parliament. Not all the background material contained in the reports is receiving the attention it may merit for use in other contexts.

The present study is seeking to increase the supply of earlier collected data, so they can be reused, evaluated and applied in new and divergent ways. We are proposing a system for factual data retrieval from government research reports in the Swedish public policy-making system. Tentatively, a data bank has been established of the contents of the Government Committee Reports (Statens Offentliga Utredningar), the SOU-series wherein also numerical data as expressions of social indicators can be found.

A report by a government study committee, including a number of recommendations to the Government, is a link between the demand for reforms and the final legislation. The Swedish parliamentarism has a rather unique characteristic in the number of such study committees which are established upon specific directives by the Government that lays down the goals and the general frame of reference. When the Government is going to take action on an issue and make proposals for new legislation to the Parliament, it relies on the fact-finding activity of especially established Government committees. Their reports together with recommendations are often included or referred to in the bills. However, before the bill is formulated, the Government sends over the report to various government bodies, agencies, organisations and others who may be concerned. They have to review the recommendations and propose changes.

After that review the Government formulates its standpoint on an issue in the form of a bill to the Parliament. The Parliament also can initiate a request to the Government to establish a committee for a specific purpose. The Government itself can either elect the individuals who are going to serve in the committee, or it can commission one of the ministers to choose the committee membership. According to Swedish legal language, the word 'kommitté' (committee) is used in the former case, while in the latter the word 'kommission' (commission) generally denotes the body. In the following all these bodies will be called 'committees'.

The committees should explore and analyse all facts of importance, so they can propose recommendations for new legislation. Their main purpose is to serve as a basis for the decision-making procedure between the Government and the Parliament. The committee activity is recognized as an essential precondition for good public policy-making. As a rule the committees submit extraordinary comprehensive and detailed reports, often also of high scientific quality. In general, their reasoning has to be based upon literal and descriptive arguments in order to disclose the present state of affairs in society, and that is especially true when a committee cannot arrive at clearly defined concepts of accuracy, precision and reliability.

However, often the committee reports contain a great many numerical data, in many cases these are scattered throughout the text and difficult to find. If the committee needs to make use of statistical data collection and processing in order to illustrate the situation,great financial resources are put to its disposal. The data are presented in numerical tables and graphs, so they can be evaluated and appraised both with regard to quality and quantity. There are great differences in the kind of data contained in various committee reports. Despite that fact, they perform essentially the same function, namely making it possible to establish a national policy and a new legislation. During its lifetime many committees collect and evaluate vast amounts of data which otherwise would not have come to surface.

Most of the government committees publish their reports in printed form in the numbered series, called 'Statens Offentliga Utredningar', or for short, the 'SOU-series'. Nowadays, almost a hundred numbers annually will appear in the series. The total cost for the government committees in the fiscal year 1979/80 amounted to about 200 mill. Sw. Cr. (=$ 40 M.). The printing costs alone run up to 9.2 mill. Cr.

The factual data in the committee reports have often been acquired by the strenuous efforts of social scientists, civil servants, experts, or research workers. However, many times their efforts are forgotten as soon as their results have been communicated through the publication of the report. Often their data have been evaluated and scrutinized, why it would be realistic to say that the data constitute a reliable information source. In principle, through the publication chain these evaluated data should be freely accessible so they could be applied also in areas outside the realm of the committee. But, too often the data are buried in the publication, and usage of the data cannot be expected because there are no direct mechanisms to access them except by browsing through hundreds of the reports.

The availability of computer operations give us the possibility to establish systems which can increase the capability to retrieve information that otherwise would be lost. The capability to identify and deal with needed information in the form of data serves as a social intelligence function in society. The access to such information through distributed information processing systems and networks is a new approach, and we can talk about 'distributed intelligence' (Dedijer 81), in the service of central and local decision-makers.

However, there are many barriers hindering a more wide use of data contained in the committee reports. One such barrier is the access-to-data problem, and in the following we will discuss how to come to grips with that problem.

Data collection and presentation

One motive for building up a specific data base from the data found in government committee reports is that the data collected by one committee should be accessible and used also by others. Usually, the data are generated and evaluated by a committee which has been set up with a specific frame of reference, and the members have been chosen accordingly, because they are familiar with that frame. But the data which have been collected and evaluated within that frame could also be needed in other subject-areas. With regard to the heavy costs for gathering the data in the first place, due attention should be paid to the need for data also in other fields. The data base presented in this paper is a means for access to data as needs arise in other subject-areas.

In the first place the data base is intended for the Government in order to make it more information efficient. However, users may even be scientists, research workers in universities and institutions, employees in government bodies, local communities and private enterprises who work on mission-oriented or problem-oriented research. In addition, technical and business consultants, school teachers, and the intelligent layman may have need for these data.

The committees try to arrive at accurate information about the current state of the problem under study, so their conclusions and recommendations can serve the Government and guide the national policy-making. It is significant that the compilations and analyses of data by a committee have as purpose to serve the decisions of the Government and the Parliament but the results may be of interest also to others. The sales promotion catalogue for the SOU-series mentions that the Government Committee Reports could be of interest also for

-those who are studying social phenomena and problems

-those who actively participate in politics

-civil servants in state government

-decision-makers in business and industry

Since the data published in the Government Committee Reports meet certain standards for accuracy and dependability, it is obvious that these data could be used more widely. Many of the committee members are specialists on various social issues. Among them are subject or trade experts, scientists, representatives of various organisations or ideologies, and as the subject areas of social issues often overlap, data collection in one field could be of interest in another.

However, there are many barriers that hinder others to rely upon data that have once been collected and evaluated. One such barrier is the access-to-data problem. The hundreds of reports published annually are independent items. There are no cross-references. The compilations and analyses of data refer to carefully chosen subject areas, mostly under one ministry. Even if subject areas may overlap, the data of one committee can have been treated differently from another committee, because the committees are handled by different ministries.

We can illustrate the access-to-data problem by the following. The bibliographic references of the committee reports are entered continuously into the Swedish national bibliography, and at the same time into the Swedish on-line bibliographic system, LIBRIS. However, the committee reports can only be searched by committee name, number or title of the report. There, are no keys for content indicators such as keywords, subject areas, etc. On the other hand, each report contains a chronological and systematic list of earlier reports issued that same year. The systematic list is arranged according to the ministries under which the committees adhere. Over a dozen ministries issue reports.

Each report has usually a short title, sometimes of ambiguous wording in order to attract interest, but not very helpful to the reader. As example can be mentioned a sloganlike, strained title such as 'Ren tur' (SOU 1979:44) having the double meaning of 'Pure Chance' or 'Clean Cruise'. Only through a subtitle is indicated that the latter meaning is the correct one, namely 'A Programme for Environment-Safe Sea-Transports'. From the subtitle that is not searchable in LIBRIS, we can infer that the report contains data about the environment. It is, however, difficult to imagine from the titles that environmental data also can be found in reports with titles such as 'Children Accidents' or in a report the title of which should read in English, 'Disabled - Integrated - Normalized - Evaluated' (SOU 1980:34). If we should rely upon title searches in cases such as these, LIBRIS is not very helpful.

The reports have rather detailed 'List of Contents' and 'List of Tables', but they lack subject index. A decision to include subject index would be a costly business and would hit upon the problem of index semantics. The indexing language would set up a new barrier for the user, because his query language may not necessarily conform with the index terms assigned by an indexer. So, a subject index does not solve the access-to-data problem.

However, if we stay on the level where we do not provide standalone information, but still references to where data can be found, we can direct the user to the primary source of data through a system of natural text searching by using the captions of the tables and graphs in the reports.

Often the captions are and should be more self-explanatory than some fancy committee report titles.

There are many kinds of data. Computer scientists talk about data banks. Information scientists talk about data bases. Economists talk about productivity data, population data, etc. Scientists and engineers talk about scientific data as results of scientific observations and measurements. In modern society, politicians and decision-makers are confronted with a whole range of, for instance, pollution problems, and the data relevant to air quality will form a long list. However, pollution problems are just part of environmental problems for which solutions additional data are needed, e.g. data to design monitoring equipment for air and water pollutants, for research on better internal combustion engines, for better efficiency and less contamination in industrial chemical plants, for finding usage for waste-products, etc.

Quantitative data as measures of quantities expressed in terms of more or less well-defined units are presented in the tables and graphs of the reports. The access keys in this project to such data are the natural text found in the captions on the tables and graphs.

Qualitative data, taken broadly, may include statements concerning the objects under study by the committee, and these statements may often be equivalent to the knowledge the committee has arrived at. The access to this knowledge can be made through searching the list of contents of the reports. Every part or paragraph referred to in the contents list, provided it deals with at least five pages, is searchable from the heading as given in the contents list. Also in this case natural text searching is performed.

The inclusion of graphs depends upon the fact that data often are obtained in graphical form when the quantity measured depends on a parameter. The parameter may be spatial distances, time, social phenomena, etc. In other cases the graphs are constructed in order to help users grasp the masses of data collected by visual perception. Charts and maps also belong to this category. Organigrams are often used to visualize the government hierarchy and the various executive branches. The method used here is to describe the graphs in the same way as tables by the captions, and the search follows the natural text.

The SAP-indexing method

The purpose of this indexing method is to produce subject descriptions of books taken from the books themselves. By selecting terms and phrases from the contents page and/or the book's index, it has been possible to create a file in which every word and 'word string' is searchable in the free text mode. This permits searching of not only a broad subject area, but even specific aspects of a book's treatment of that subject. Using the table of contents and the index, one selects terms and phrases which point to the significant subjects discussed in a book. This information is arranged and processed so that one locates not only the book, but the specific page(s) within the book on which that subject is discussed.

In an on-line system then, the answers to queries will be lists of chapters, parts, sections or paragraphs to books, where the references are arranged in descending order according to the number of pages dealing with the subject asked for.

Contrary to other methods for deep indexing, this one does not require any special subject knowledge on the part of the indexer. When it comes to selecting words from indexes and contents pages, rules and guidelines have been devised to ensure that both significant subjects and useful search terms will be extracted from the book. The criteria for selection are easy to understand and easy to apply, though they are controlled by quantitative standards: the number of selection entries, page ranges and the quota formula.

Although the results of the SAP test data base are satisfactory, further investigation into this form of subject description is necessary. The length of the description and the rules for selection need to be reviewed critically. In its present form the quota system is difficult to apply and probably inadequate.

As subject descriptions are merely extracts from each book, some effort must be made to improve the contents page and indexes in books so that these extracts will be more content-bearing. The publishers should be more aware of the fact that the key to greater access to book collections rests partly with them. The books they produce provide the subject descriptions we can use for our augmented on-line catalogs of the future. The standards for indexing should be seen as the means to insure that the index will allow the reader the most precise and effective access to a book's contents.

Characteristics of the data base

The identification of data requires the detailed description of the properties, substances, entities, phenomena and measuring conditions. These descriptions can be more precise than the bibliographical descriptions of a committee report. Classification or indexing in the conventional way is not necessary, since a powerful software system for information retrieval can easily deal with the natural text of a description. The data in the committee reports are described by the text of the tables and graphs, and that text serves as input in the data base. With regard to flagging the data, the data base distinguishes between three categories, namely

1. contents terms - CT - words in the text taken out of the contents list

2. table terms - TT - words taken from the captions of the tables
3. diagrams and graphs terms - DT - words taken from the captions of graphs

The present potentials of retrieval systems have made it possible to refine the retrieval process by such a more heavy use of free text than what was possible in earlier manual systems. The manual systems had no equivalent to the searching of such new elements as captions of tables and graphs. The size of the data base when this is written, amounts to 120 government committee reports, covering those published during the last year and those of the current year. Besides the flagging of the three categories above, there is no further classification. Because of homonymes that may result in false drops. However, since the dialogue with the computer data base is carried on in real time, such occurrences are easily discarded.

We do believe that the captions to tables and graphs are reliable text strings that can be searched, because the authors and editors are at present more aware of the international recommendations made (CODATA 73) that figures and tables shall be logically complete units, independent of the main text. The caption should explain the table or figure, and it is often recommended that particular care should be taken in labelling numerical data that may be cited independently.

Input equipment and procedure

Information has been put into computer-readable form through the keyboard of a Texas Instrument Silent 700 onto the tape of a microprocessor, MFE 5000 Buffered Data Terminal. The data have then been sent over the telephone line in batch processing mode to a DEC 10 computer in Gothenburgh. The microprocessor is used for editing purpose, since spelling can be corrected, missing words inserted or superfluous words deleted, before entering the DEC 10.

The bibliographic description follows the convention of AACR.

The input of the contents list follows the rules set forth in the SAP-project at the Syracuse University under the guidance of prof. P. Atherton-Cochran, namely 'Subject Description of Books. A Manual for Augmenting Subject Description in Library Catalogs. Ed. by B. Settel. Syracuse, 1977'. According to that only text strings out of the contents list that refer to at least five pages are candidates for inclusion in the data base. Irene Wormell, the co-author of this report, has made a Swedish translation and guide-book of the SAP-technique that has been available during the project, namely 'Handbok för SAP-indexering, Lund, 1979. (LULID5/LIBR-1600)'.

It was found expedient to make photocopies of the contents list and let a documentalist mark the headings which should be included in the data base. Additionally to the SAP-technique, the data base has been augmented by the captions to all tables and graphs in the committee reports. This way of deep-indexing of books is such a straightforward procedure that it was done by a clerical effort alone.

The intention is to cover ten years of publication. On the average, the 120 reports that have been included already, generated records of 4.000 characters length.

The software - the 3RIP system

Government funds for research has made it possible to construct an interactive text data base system that has been labelled 3RIP. That system has been implemented for testing purpose in a DEC 10 computer at Medicindata, Göteborg University. The Medicindata is the node and the monitoring agency of the SCANNET, the packet-switched network between the Scandinavian countries. The data base, called 'SOU', is placed on the DEC 10 using 3RIP software.

3RIP consists of:

- a retrieval system used to locate the records in the data base and display them in various ways,

- a file handler used to read records from the data base and to write records onto files,

- a text editor used to create and modify records.

The records may be of any length as long as they can be read into the core in their entirety. That has been an advantage in our case, since some of the records exceed several thousands of charachters.

3RIP can handle the contents of up to 30 different search fields. In our case six fields have been set up as searchable, namely: titles = TI, words in the contents list = CT, corporate author = AU, report numbers = NR, captions to tables = TT, and captions to graphs = DT. A search order such as:

AU = BUDGETDEPARTEMENTET AND TT = RATIONAL/

will identify records where the Department of Budget has issued a report about rationalisation. Left- and righthand truncation is possible, as the example shows for the word RATIONAL/.

The output

The output from the searches comes on screen or typewriter terminals. In the search file two output formats are available, for instance the above command AU = BUDGETDEPARTEMENTET AND TT = RATIONAL/ will result in

```
FORMAT 1
    (ISBN)        91-38-05115-X
                  SOU 1979:72

    (AU)          BUDGETDEPARTEMENTET

    (TI)          RATIONALISERING OCH ADB I STATSFÖRVALTNINGEN.
                  BETÄNKANDE AV UTREDNINGEN OM CENTRAL RATIONA-
                  LISERING OCH ADB

FORMAT 2

                  91-38-05115-X
                  SOU 1979:72
```

BUDGETDEPARTEMENTET
RATIONALISERING OCH ADB I STATSFÖRVALTNINGEN.
BETÄNKANDE AV UTREDNINGEN OM CENTRAL RATIONALISERING OCH ADB

TT
(CAPTION of table) PROGRAMMET RATIONALISERING UPPDELAT I DELPROGRAM (44)

As can be seen the Format 1 gives the bibliographic reference, while Format 2 also focuses on the text which has caused the hit. In this case we hit upon the word 'RATIONALISERING' as contained in a caption to a table found on page 44 in the report.

Conclusions

As has been shown above, the direct access storage technology permits a variety of possible file organizations for adaption to the contents and uses of a data base. If we define an information processing system as the use of logic and technology to provide an existing service more efficiently or a new service that was not previously economically feasible (Harden 80), our data base is a good example of that. The service besides the LIBRIS system has been improved, and the records have been augmented to such an extent that we can talk about a new service to the users.

The present project may be seen as an effort to automate a simple routine of clerical nature, namely to create a data base of government research reports to moderate costs, and still the hope is that its usage can improve the productivity of the politicians and decision-makers. The captions to the tables and graphs refer to information collected objectively and systematically. The references to parts or paragraphs in the reports taken out of the contents lists refer to the interpretation of the information in response to the frame of reference of each committee. In the former case, we are dealing with factual information of objective nature, in the latter with subjective interpretations or even best guesses. The data base, just as the reports, is in the Swedish language. Occasionally, however, a report may have an English summary.

Acknowledgement

The idea of title augmentation by extracts from the contents lists stems from prof. Pauline Atherton-Cochran to whom the authors are indebted. The innovative way of using the SAP technique for captions to tables and graphs stems from discussions with the treasurer of CODATA, Dr. D.G. Watson, Cambrige Chemical Laboratory. The Delegation for Scientific and Technical Information - DFI - has given the financial support and endorsement to the project. The staff of the Medicindata have been most helpful to adjust and implement the 3RIP software to fill our demands for specific output formats. In a broader context the SOU data base is part of a program to review the 3RIP software in various environments.

References

1. P.Atherton et al., 'Books are for Use.' Final Report of the Subject Access Procedures for Augmenting Subject Descriptions in Library Catalogs. Syracuse University, School of Information Studies, 1978.

2. CODATA. - 'Guide for the Presentation in the Primary Literature of Numerical Data Derived from Experiments.' - CODATA Bul. 9, 1973.

3. S. Dedijer, 'Distributed Information Processing - Distributed Decision Systems - Distributed Intelligence Systems.' Research Paper. Lund, 1981.

4. Y. Dror, *Public Policymaking Reexamined*. New York, Intext 1968. 378 p.

5. R.M. Harden, 'Management of Information in the Executive Office of the President.' Washington, 1979. 18 p.

6. R.M. Harden, 'Toward an Information Efficient Executive Office of the President.' Washington, Office of Administration, 1980. 69 p.

7. J.E. Kasputys, 'Private and Public Sector Responsibility for the Collection, Distribution and Analysis of Statistical Data.' - *IASSIST Newsletter* 4, 1980, 3/4 pp. 69-72.

8. B. Settel, (ed.), 'Subject Descriptions of Books.' A Manual of Procedures for Augmenting Subject Descriptions in Library Catalogs. Syracuse University, School of Information Studies, 1977.

9. B.V. Tell, 'Retrieval Efficiency from Titles and the Cost of Indexing.' - Inform. Stor. Retr. 7, 1971, pp. 241-43.

10. I. Wormell, 'SAP - A New Way to Produce Subject Descriptions of Books.' - J Inform. Sc. 3, 1981, pp. 39-43.

11. I. Wormell, 'Handbok för SAP-indexering'. Lund 1979. (LULID5 /LIBR-1600.)

THE WASTEFUL USE OF EMPIRICAL STATISTICAL INFORMATION IN INFORMATION RESEARCH

Bertram C. Brookes

1. Natural and artificial sciences

Herbert Simon, an American economist who was a Nobel laureate in 1978, has been much concerned with the problems of quantification in economics. An early book of his, The Sciences of the Artificial, has recently been re-published in an expanded second edition. In this book, Simon distinguishes those sciences in which the entities studied are artefacts from those in which they are natural and discusses differences that this distinction makes when quantified effects are analysed.

With money as an object of study as well as the main mediating entity, economics is the archetypal science of the artificial. But, taking a fundamental view, I see information science as a potential natural science. At least its mediating entity, information, is natural and so too are the information processes which occur in our neural systems. But the documents, libraries, databases, computers,.....with which we are professionally concerned, are obviously artefacts. So information science in its present state of development and as currently practised has to be regarded as Simon regarded the engineering sciences - as one of the sciences of the artificial.

Simon warns that in the sciences of the artificial phenomena do not wholly conform with the inexorable laws of science. As an economist he is particularly concerned with optimisation techniques and introduces the term satisfice to express the fact that we often have to be satisfied with what we can get rather than with what an ideal theory predicts.

In information research, for example, it is fashionable to set up theoretical models of, for example, retrieval system performance, to apply mathematical or statistical ideas to these models and thus to reach theoretical conclusions. This type of research is common in other sciences too, particularly in engineering. After long experience, engineers have learned that no theoretical model about engineering artefacts can be taken seriously until it has been validated by tests in the real world. In engineering, in which the artefacts are physical objects, any working artefact is a compromise between the theoretical ideal and the contingently possible. In the cognitive world of information science theoretical ideals and everyday practicalities are even more divergent than those of the physical world. Models must be tested.

2. Hierarchical modes of thought

Herbert Simon also stresses the importance to analytical studies of hierarchical frameworks of thought. Such frameworks underlie library classifications and information retrieval processes, so the idea of switching the focus of attention from one cognitive level to another is familiar to us.

Hierarchical frameworks are used to reduce highly complex systems to a discernable order. When the number of entities in a field of thought increases beyond some surprisingly small number, 8 or 10 maybe, we usually begin to try to group them into classes and to name the new classes. Thus several entities which demand individual identities at one cognitive level may be compacted into a single class, a single member entity of the next higher level of generality. Discussion of the highly complex system becomes simpler when the number of inter-related entities is reduced. So hierarchical frameworks can lead to economies of information processing and, thus, to economies of thought. And these economies are measureable.

If n identifiable entities are aggregated into m classes or groups we then have two levels of an hierarchical structure. The information in the original level is proportional to $\log n$ and in the more general level to $\log m$. For $n = 100$ and $m = 10$, $\log n/\log m$

= 2. The 10-fold reduction in the number of entities leads to a 2-fold reduction in the information that has to be handled. The classification itself has also added new information - half as much again to the initial logn.

When argument becomes difficult because of the complexity of the topic, attempts may be made by one or more disputants to raise the arguments to a more general level. The advantage of this technique is that by reducing the number of distinguishable entities one also reduces the complexity of the argument. However, the argument may not be settled by such a move because the world around us is so complex that no one hierarchical framework serves all purposes - the world of thought is polyhierarchical. So different disputants may try to impose different frameworks on the argument, and argument continues.

An alternative way of trying to resolve complex issues is to shift the argument to more specific levels. But this move involves a search for more information relevant to the issues. This is the traditional way of science. It calls for further observation, experiment, information, and we now have computers to help in the analysis.

As information scientists, we have a special responsibility to use our stores and other sources of information knowingly and with an understanding of information economy.

3. The individualities of information work

The characteristic of information science which distinguishes it from all other social sciences is that it is fundamentally concerned with individualities of both users and of the information sources they seek. At the macro-level, a data-base such as MEDLARS can be regarded as a class of documents collected to serve the interests of the medical profession. But the basic intriguing problem of information work is to discover how to help the individual user to find the particular bit of information he seeks from within the total collection. The answers we seek to this problem are not going to be found by theoretical analysis at the macro-level of classes, which is the cognitive level at which conventional statistics operates. At this level, the individual user and his individual problem are lost to view in the averages and totalities of data-base operations. The theoretical development of information science will require quantitative techniques which recognise individuality, as do, for example, ranking techniques.

Human activities can be studied at either (a) Level G - the level of generality at which humans are grouped into classes, or at (b) Level I - the level at which individual performance is studied. Because Level G is typical of the physical sciences (which have no Level I) it is often assumed that analysis at this level offers the only truly scientific approach. A further attraction of working at Level G is that it is abundantly provided with the analytical techniques of conventional statistics which are now mechanised; data can be punched into the computer and be completely analysed by the standard packages - but only, of course, at Level G. (The conventional statistics of ranked data, also available in package form, is erroneous).

If the basic problems of information science lie at Level I and the proffered solutions lie at Level G, then problems and solutions will continue to pass each other by.

4. Statistical information in a library usage survey

A simple but typical example of a project in the information field is given by a survey of library usage. The purpose of the survey might be to discover the basic patterns of the usage, basically: Who are the users and what is the pattern of their borrowings? and/or Which are the books and what is the pattern of their borrowings? As both queries give rise to similar statistical considerations, I shall focus only on the user patterns.

I shall assume that the users could be named individually though I appreciate that such naming could well be regarded as an objectionable intrusion into private transactions. I would therefore accept a coded form which, while admitting the individuality of the user, hides his actual identity. Books too have their unique bibliographic descriptions but for survey purposes books too could be represented by some identifying number.

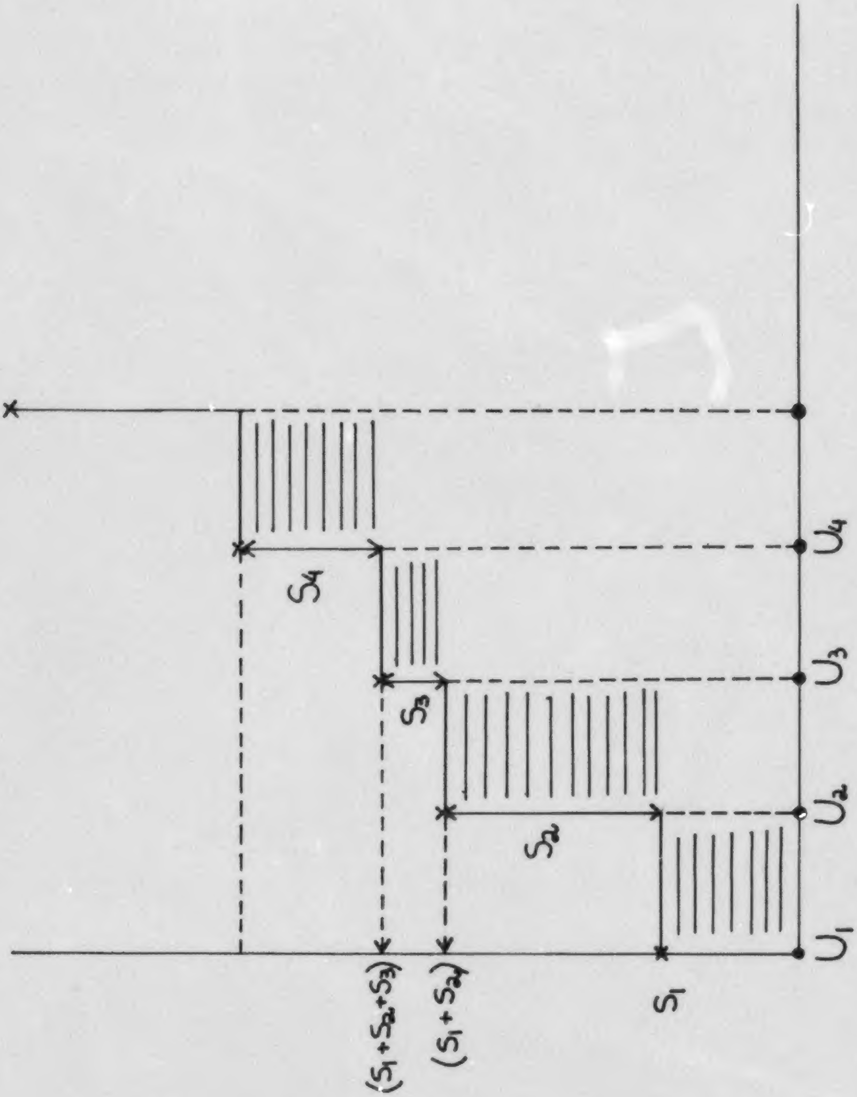
So the raw data of the survey would consist of a list of transactions in which, in principle, the name of a borrower (coded) is linked to the title of a book (also coded). Ideally, it should be possible to record all such transactions automatically and so have, at the end of the survey period, a time-sequence of transactions recorded as binary entries in which a user's name is associated with a book title. This list, which I will call File A, constitutes the raw data of the survey.

Suppose the number of transactions in File A is P and that the total number of users named is N. To discover any regular patterns in the transactions, the survey would have to continue for a period which would permit the most active users to borrow several books. So we would expect P to be greater, perhaps much greater, than N.

The first step in any statistical analysis is to collect the numbers from the data. In this case, File A could be re-arranged: as each user name appears on the file, the rest of the file is scanned to count all subsequent transactions in which that user is involved and to bring forward these entries so that we see against each user name the files of all the books he has borrowed. Let us call this new file, File B.

No information presented by File A has been lost by re-arranging the data to form File B; in fact, some numbers have been added. If the user U_r has borrowed several books, say S_r, let us call S_r his 'score'. Then the sum of all such scores, S_r should equal the number of transactions, P.

A graph could now be drawn to depict the information of File B in the form of two arrays. One array, the list of users, is marked off along the x-axis of the graph. The second array, that of the numbers of books borrowed, or the scores, are marked off cumulatively up the y-axis. Thus S_1 is plotted against U_1, $(S_1 + S_2)$ is plotted against U_2, $(S_1 + S_2 + S_3)$ against U_3, and so on. The advantage of using the cumulated scores is that such a graph could be annotated to list the titles of all books borrowed by each user, user names along the x-axis and book titles, P of them, up the y-axis.


(S1+S2+S3)
(S1+S2)
S1
S2
S3
S4
U1
U2
U3
U4

Such a graph is of no theoretical interest but I mention it to emphasize that it completely and very directly depicts the actual usage of the library as represented by the raw data. If the Files A and B were lost, the raw data could be completely reconstituted from the annotated graph.

This graph of File B is otherwise of no interest, but by ranking the users of File B according to their scores to produce a third file, File C, a much more revealing graph could be drawn. Again, this graph presents all the data inherent in File A - no information is yet lost. But how much statistical information have we available in Files A, B and C?

We apply the usual argument from information theory. Here we have P events distributed over N users, with P greater than N. The first event, the first book borrowed, could have been borrowed by any one of the N users, so there were N possible distributions of the first book. Unless there are borrowing restrictions on users, the second book could also have been borrowed by any one of the N users. There are thus N^2 ways in which the first two books could have been distributed. The argument continues similarly for all P events. We conclude that the number of possible distributions of P events over N entities is N^P and of this large number our survey captured just one.

The measure of the statistical information thus captured is given by the logarithm of the number of possibilities. So the statistical information I(A) presented by File A is given by

$$I(A) = \log N^P = P \log N \qquad (1)$$

5. The conventional statistical approach

If the raw data of File A were presented to a conventional statistician, his first action would be to re-cast the data into the form of a frequency distribution. He would ask: How many users borrowed only one book? How many two books? How many three? and so on. And if some of these numbers became small, say less than 5, he would probably combine classes so that overall he had about 10 classes with their respective frequencies.

Note that this distribution is one in which the N users are distributed, not the P events, and over 10 or so classes. If the number of classes is C, then by the previous argument, the number of possible distributions is C^N. And so the statistical information inherent in the frequency distribution is I(F) where:-

$$I(F) = \log C^N = N \log C \quad \ldots\ldots\ldots\ldots\ldots(2)$$

Comparing this result with (1), and remembering that P is greater than N and that N is greater than C, it is evident that

$$P \log N > N \log C \quad \ldots\ldots\ldots\ldots\ldots(3)$$

With $P = 500$, $N = 100$, $C = 10$, $I(A) : I(F) = 500 \log 100 : 100 \log 10 = 10:1$. and with $P = 5000$, $N = 1000$, $C = 10$, $I(A) : I(F) = 15:1$.

The fact that something has been lost can be verified by trying to reconstruct File A from the frequency distribution alone - it cannot be done.

5. Concluding comments

If the basic phenomenon of information science is that of an individual user seeking some particular item of information, it seems to me that the statistical calculus of classes is too crude an instrument to be useful. If the analyst begins by discarding 90% or more of the statistical information provided by the empirical study, it is not surprising that theoretical development in information science is rather slow.

Reference

Simon, Herbert A. The Sciences of the Artificial. 2nd ed. MIT Boston, 1981.

RESEARCH ON THE NEEDS AND USES OF SCIENTIFIC AND TECHNICAL INFORMATION. RESEARCH AREAS AND PROBLEMS.

Lars Höglund & Olle Persson

Introduction

Information supply and communication processes are fundamental to science and technology in about the same way as energy and matter are essential to material production. Simply stated, the task of research and development is to produce knowledge using earlier knowledge as a point of departure. To accomplish this task, scientists and engineers are dependent upon the effectiveness of various communication systems.

The interest in improving the communication of scientific and technical information has, in Sweden and other countries, resulted in the establishment of central authorities responsible for R&D-activities in the field of Library and information science (LIS). The field of LIS includes a broad spectrum of problems related to the communication process, which have to be approached from several disciplines of technology and social sciences. In this report we shall pay special attention to the problems of information needs and the uses of scientific and technical information, and describe important areas in need of research and development.

The fact that Swedish R&D is a small fraction of the world's total R&D suggests that the new knowledge we can produce in Sweden is very limited compared with the existing body of knowledge and on-going research in other countries. From the point of view of national science policy, import of knowledge from abroad is of utmost importance.

Firstly, information use is an important basis for deciding if a certain kind of research should be supported and how it should be conducted.

Secondly, information use is a major prerequisite for the research work, which to a large extent consists of collection, interpretation, transformation and dissemination of information.

Thirdly, information use is essential for the maintenance and development of professional compentence in society at large. Besides R&D-professionals, there are many other groups that need information about science and technology in their own practices, as well as for the democratic control of research and development.

In the field of LIS most activities have been directed towards improving the physical availability of information by utilizing computer and telecommunication technology. The development of technical facilities and marketing of new information services often includes assumptions that R&D-work is a rational, cumulative process of knowledge production, where new results are added to previous knowledge. The awareness among users of the value and availability of information services has often been taken for granted. These assumptions can be seriously doubted since several user studies have demonstrated a considerable lack of knowledge about information sources and services among R&D-personnel. Furthermore, a number of social, economic, and psychological barriers restrict both motivation for information use and the possibilities for information search and aquisition. The non-use of information sources and systems does not necessarily depend on the properties of formal information systems. Instead, to a large extent, such problems might be attributed to conditions related to the users and their environment. Consequently, research should be devoted to find ways to overcome information barriers and to stimulate motivation for information use. To start with, we can distinguish between obstacles related to the physical availability of information, the form and content of information, and the user's capability to search, understand, and make use of information. Besides the qualities related to physical availability, we also have to study the more 'soft' and 'vague' aspects behind the communication behaviour.

The field of information science covers a whole communication process where each step can be a subject of study. This implies that information science is an interdisciplinary field which is related to both science, technology, and social science, as well as the humanities.

In Sweden information science has developed from three fields of research and practice: library science and practice, computer science and technology, and social and behavioral science. Within the social sciences, the sociology of science and communication research are two areas of special importance for the understanding of scientific communication as a human activity.

The purpose of this paper is to review important research areas related to needs, uses, and dissemination of scientific and technical information from the perspective of the social sciences.

To review important areas of research in a new field is not an easy task and calls for compromises between the use of established approaches and new directions for research, between short term requirements and more basic or long-term research programmes. Of course, there are no such things as objective reviews or research approaches that are the only possible roads to knowledge. Considering this, and the space limitations for this paper, we have chosen to treat a number of topics without explicit statements on how questions should be formulated or data collected. Such problems are left to the fantasy and eager curiosity of individual scientists and experts.

Problems of integrating the I&D functions in the user environment

University libraries and libraries in industrial organizations often lack integration in policymaking and administration.

The awareness among authorities responsible for R&D of the problems that confront libraries and information centers seems to be a strategic factor if there is to be any fundamental change of the information and documentation functions. This is not only determined by the fact that the funding of such functions has to compete with other activities, but also by the need to stimulate researchers' and other professionals' motivation for information search. Information needs, as well as the conditions for information use, are fundamentally determined by the content, organization, and goals of the user's work.

One task for empirical research in this area is to compare different organizations with regard to their methods of handling communication flows and their organization of information functions.

On a national level, information supply is not often discussed in close relation to science policy. However it is well known that information and documentation workers at all levels often express a need for better coordination and integration with the production, research or decision-making processes they are serving.

I&D-policy can partly be looked upon as a special case of R&D-policy, on the level of organizations, as well as on a national and international level. How different policy issues are formulated and approached on these levels should be studied as a basis for the development and evaluation of information strategies. To start with, it is desirable to review current policies and programmes related to the field of LIS using a comparative approach.

User groups - a need for basic statistics

User groups for scientific and technical information can be defined in several ways. One way is to separate users according to the type of output they produce. For example, researchers use information in order to produce new information, practitioners produce commodities and services, decision makers need information to arrive at decisions, and teachers or journalists use information in order to disseminate it to students or others.

Other classifications of users can be made with reference to subject area, the degree of basic or applied research, type of organization, etc. User groups can also be distinguished according to their actual demand for various information services.

More complex classifications are possible if two or more dimensions are combined.

The research activities that generate information needs are sparely described in official statistics. The number and type of researchers and other professionals engaged in R&D is difficult to estimate, and the user groups in universities are difficult to compare with industrial users, even in bare numbers.

The need for better R&D-statistics has been recognized now and then during the past decade. The choice of R&D-indicators is not easy, and has been discussed in several countries and OECD. Especially when it comes to measures of output and productivity, disagreement about the value of such measures or their validity can add to the difficulties. However, for the planning of I&D, rather simple input-oriented measures would create a basis for the definition and estimation of user groups, their size, and development over time.

Another approach, which is a necessary complement to R&D-statistics, is to describe the actual users of information systems and services. This can be compared with the potential user groups, thus, indicating the degree of use and non-use, as well as user groups not covered by the statistics.

A third step in this line of reasoning is to start from the supply of information. Statistics covering the libraries, information systems, and services are however, also scarce. Such statistics should be possible to match with the statistics covering user groups, in order to estimate the degree of existing needs that are possible to cover with the present supply of information in libraries and information centers.

Three kinds of basic statistics are needed tc get a starting point for a general planning of I&D-activities at the national level: R&D-statistics to define the structure of potential user groups, statistics about the actual demand for information services, and the existing supply of information resources available.

Information needs

The general aim of I&D is to satisfy information needs among various user groups. However, it is very difficult to define information needs.

Firstly, we can distinguish between 'objective' and 'subjective' needs, i.e. the needs that can be estimated from the character of the problems that the users are confronted with and the needs that the users themselves perceive. To what extent these two aspects of needs really coincide is not known.

A more simple approach is to rely more on actual demand for information and use this as a crude indicator of information needs. Perceived satisfaction with information services used is another kind of data which indicates the ability of information sources to satisfy needs.

Still another approach, which does not depend on the users previous knowledge about the existing supply of information sources, is the experimental approach. User groups can be offered new information services and the outcome can be studied in terms of use, satisfaction, and perhaps effects on the ability to solve problems. Experiments can also be used as a tool for technical assessment of information systems, indicating what aspects of a system that have to be modified and in what way information systems should best be administrated, the need for user education, etc. In general, the experimental approach will be more and more important since new information services are continously being developed. The information market in science and technology has been changing drastically during the last ten years and it is not probable that the user can express needs regarding new products without having the possibility to test them.

Information use

To describe the use of various information sources among scientists and engineers has been the purpose of many so-called user studies. In order to measure information use, several methods have been employed. In surveys it is common to ask for the frequency of use and the perceived value of information sources. Another method to estimate use is the critical incident technique, where questions are asked about the source of information last used or the last incident when valuable information was received. In other user studies specific information systems are analyzed, for example, requests at libraries, evaluations of information retrieval systems, and studies of readers/subscribers of journals, etc.

Data from user surveys could be utilized in planning I&D-services; what services should be developed, expected demand, specific preferences, the need for user education, back-up service, etc.

Descriptive data on information use are essential for I&D-planning. However, user research should also be directed towards finding causal models capable of explaining preferences and patterns of use. In the research literature on information use, a few explanatory models have been proposed simply by listing factors that are related to the use of information. Such factors are, for example, the degree of basic research, subject area, the organization of research work, the social position of users, their work tasks, and individual properties and characteristics of the information environment. The relationships between such factors are complex and have only partly been analyzed. In order to stimulate information use it is important to have a more thorough understanding of the determinants underlying information use, if adequate measures are to be found.

Most user studies have stopped at the level of exposure to information sources. However, the concept of information use includes several other aspects of interest. The information has to be searched, read, interpreted, and evaluated. When using reference sources, documents have to be acquired which may involve special problems in terms of costs, delivery, and time. The use of information after acquisition has not been the subject of study and the knowledge of how information is used in the research process is very limited.

A special problem area is the effects of information overload and how to achieve a balance between the consumption of information, cognitive capacity, optimal stress-levels, and problem-solving capacity. In order to study such questions and others related to the actual communication behavior within the research act, there is a need for more detailed case studies, for example, including participant observation. Such research projects would supplement the survey studies proposed above.

Time and information use

Information use is in many respects related to time. First of all, information use is a very time-consuming activity in R&D. Studies of time spent communicating in R&D-laboratories have shown that between 50 and 80 percent of the working day is devoted to communication activities, where reading of literature takes about 20 percent. These figures imply that there is little time left for an increase in information use. Scientists, engineers, and other professionals in R&D have to balance information use with other activities. Therefore, the free time capacity for information use is an important factor to consider in the planning of information services.

Since information activities are so time-consuming, productivity ought to be stimulated if the time spent on search, aquisition, and handling of literature can be reduced. Several user studies have demonstrated that users tend to select information sources that are easily available such as colleagues and private book shelves or files. This implies that the physical proximity of an information source is a more important criterion than the potential value. Research considering the time-reducing effect of different information systems, as well as studies of response times and users' time requirements, is lacking.

Both information needs and uses vary during the R&D-process. In the early phases when ideas are generated, information use is very intense and several sources are used. In the phase of problem solving, written sources seem to play a more important role than informal contacts. Discussions with colleagues will increase when data are gathered and when experiments or tests are made.

Of course, the preferences for information sources do not follow this somewhat simplified model in all cases. Further research on this topic is motivated by the need for the adaption of information services to the working conditions in R&D.

Another aspect of time is the age of the information used. It is known that the literature cited or ordered by scientists is much younger than the literature used by social scientists or researchers in the humanities. Accordingly, the demand for current literature will be greater in some areas than others. Knowledge of the age distributions of the information sources used can be applied to the planning of holdings in libraries or secondary services.

The measurement of time used for information activities is often used as an indication of cost, when the benefits or effectiveness of information services are studied. In such studies the time used for different activities has been investigated among R&D-professionals, as well as among librarians. Results from studies of cost-effectiveness can be used as a basis for an organization of the information work and the introduction of new information technology.

Effects of information use

The relation between information use and productivity, or problem-solving capacity, is naturally a strategic factor in discussing the value of information systems. The problem is not whether there is a relationship, but what the character and magnitude of such relationships are. A number of studies show that high performers in R&D are also frequent users of both oral and written information sources.

The effects of information systems and programs can also be analyzed at the organizational level. Comparative analyses of organizations may indicate the effects of information services in use and the possibilities to coordinate the supply of information.

Informal communication

A large portion of the information input in R&D comes from personal contacts. We already know that most ideas and innovations stem from informal sources, and that the most productive scientists and engineers have extensive personal contacts with people both within and outside their own organization. The importance of informal communication is due to the fact that not all information of relevance for users is available in written form. Personal experience and know-how is a necessary complement to printed documents, and contacts with colleagues make it possible to receive information much faster and in a more condensed form compared to the use of literature.

Informal communication has been studied in R&D-organizations and among scientists in research areas. At the level of organizations most studies try to identify people who can be assumed to play a key role in linking the internal communication network with the outside world. The way information is actually spread in internal networks, the direction of information flows, and the effects of various networks have been given less attention. Only to rely on gatekeepers and other special communicators seems to be a far too limited strategy, if the general information consciousness is to be improved within R&D-organizations.

In studies of communication structures within research fields it has been shown that the networks are built up around certain individuals and research centers. These networks are quite elite and are not open for all. The barriers to access to informal groupings is a research task that should be undertaken.

Since specialists in various field are geographically scattered, in many instances all around the earth, there is a need for communication technologies that are capable of reducing distance and time barriers. Telephone conferences, computerized teleconferences, and videomeetings are examples of media that have been paid some attention to in limited experiments.

The variability of the applications possible implies that it is difficult to generalize about the value of such media without a systematic experimental design. Taken as a whole, new telecommunication technology offers a wide range of possibilities to combine media. Studies should be made of different multimedia arrangements where users are able to switch between information channels.

Evaluation studies of new communication forms are necessary. However, such studies must rely on knowledge about human communication processes, if a deeper understanding shall be possible. Studies of the informal, face-to-face communication mentioned above, are therefore of high priority.

Communication within research areas

R&D-work can be looked upon as an integrated part of the research specialities which are continously changing - they grow, divide into subareas, decline, and eventually disappear. Social studies of the development of research areas have shown a close relationship between various stages of growth and the forms of communication. In younger areas the forms of communication are less developed and the exchange of ideas is to a large extent dependent on informal contacts between active researchers. In more established areas the formal channels are of greater importance and the information about ongoing work is available to a much larger public.

Research areas and technical specialities are international social systems. The possibilities to follow their development are therefore dependent on the contacts that users have with their colleagues abroad, as well as within their own country. These contacts are dependent upon the social position, competence, and productivity of the user - they are not available to anyone. The concept of invisible colleges indicates that the communication networks are both hard to identify and elite in nature. The international character of such networks also implies that there is a need for extensive travelling if a researcher should be able to keep up with ongoing research.

The information needs and uses of individual researchers will be partly dependent upon the social and intellectual structure, and his position in these structures. Our knowledge about how communication structures are developed in research areas is very limited and should be extended.

The capacity of formal channels to inform users about the structure of a research field might be improved by research on citation networks. Citation networks have the advantage of describing the current state and development of research. Maps of scientific specialities have been produced using citation indexes showing the most productive and influential documents and authors as well as their mutual relations.

Another task is to study the actual international contacts available to Swedish R&D-professionals. What are the economic and practical barriers towards international information exchange and what advantages do international contacts have, are questions that should be answered.

Within R&D-areas it would also be valuable to study the contacts and communication barriers between different R&D-organizations, such as industry, university, and private research institutes. The communication between such organizations, with different relations to basic research and practice, might be important for the exchange of research ideas, as well as for the rate of innovation in industry. The contacts between research and practice, between science and industry, have been discussed in Sweden, but have seldom been the subject of study.

User education and marketing information

The lack of balance that in many cases exists between the supply of information and the demand, could to a certain degree be explained by a lack of knowledge and experience of information sources. One reason for the lack of knowledge in these respects is that users have not been trained in using library facilities during their professional education.

The purpose of higher education is to prepare students for the entrance into the labor market. Qualified education should also be an innovative and dynamic resource in organizations and in society at large. In a study of students at a Swedish technical institute, questions were asked about the role of information use in education and work. It was shown that the great need for information use perceived by students, when it comes to the demands of working life, did not correspond to the demands expressed by teachers during education. This somewhat contradictory situation indicates that there has to be radical changes of the pedagogical methods in such a way that information search and use are both demanded and rewarded. In Sweden there are some experiments which try to integrate education with training in library use and information search. Experiences from such experiments should be evaluated and could be used for development of more general recommendations.

User education could be looked upon as a special form of marketing information systems. Besides training courses, there is also a need to inform users of the existence of information systems, their advantages and limitations. In some cases, information campaigns are needed to stimulate information consciousness at a more general level, both in terms of motivating users to use information services and of enabling them to express their needs of new or better information systems. However, to succeed in marketing information services, information providers have to have a more profound knowledge of the values and attitudes among users towards information use and communication. Even if the normative structure of science and technology stresses the need for open communication and rational principles for information use, there is empirical evidence indicating deviance from such norms. Competition and personal motives can be assumed to influence communication behavior.

Production and dissemination of information

Information use and production could be looked upon as two sides of the same coin. In most cases information is used in order to produce new information - and the same channels are used for both purposes. Research results are disseminated through a large number of information channels, which can create confusion from the viewpoint of users in deciding what sources to be selected on a specific topic. There is also a considerable time lag between information sources. Studies of the communication history of research results have shown that several months may pass before a result is presented in journals or other forms of publication. Before original documents are incorporated into secondary sources, many months may pass and when it comes to books, it can take several years before results are presented.

To the wide range of problems confronted within the field of I&D, one can also include problems related to the publishing of research results. In general it is something of a paradox that so much money is spent on doing R&D and so little on making the output available for a broad audience. Relevant knowledge may more or less get lost because of rather futile practical limitations, such as poorly defined target populations, bad printing, etc. Language and cultural barriers will also limit the diffusion effects of information channels.

The so called grey literature is a special problem to libraries and secondary services. A few attempts have been made to register and control the vast number of small report series, but in most instances the diffusion of reports is subject to individual decisions by authors based on a subjective knowledge of receivers. The amount of grey literature, the way it is disseminated and used, should be analyzed in more detail.

Different forms of dissemination will give various effects in terms of the number of people receiving the information, and the degree to which it is understood. Today we lack knowledge about what diffusion effects different information channels have. For example, one should study how well Swedish research results are disseminated within and outside the country. One way to do such a study is to use various forms of citation analyses - comparing citations with what is actually published.

Evaluation of information systems

Information systems can be evaluated from many perspectives. From the viewpoint of the users, the ultimate value of information systems is their capacity to satisfy information needs. The user perspective also includes the interaction with information systems, for example, the dialogue with databases, the search for literature in library catalogues, the use of bibliographies, etc. Formal information systems, such as libraries, information centers, and on-line services, can also be evaluated at an operative and administrative level in terms of both technical and economic criteria.

The concept of evaluation in information science is especially associated with information retrieval systems. Several approaches can be distinguished. The comparative approach, which becomes more and more important, tries to make systematic comparisons between the properties of many IR-systems. Historically, IR-systems were evaluated with respect to output criteria - the number of retrieved references, precision, and recall. However, these criteria only give a description of out-comes and do not give any indication of the causes of system performance. Later evaluation studies of IR-systems have extended the list of criteria in order to include the whole search process and the external factors influencing it. The comparative approach is especially relevant to decisions about what system to choose in specific situations. The other approach focuses on individual systems, and the evaluation results can be used as an integral part of system development and administration. Both of these approaches are desirable and complementary.

Evaluation studies should not be limited to computerized information systems. A lot of other information sources and services might need evaluation studies in order to find ways to improved economy and serviceability.

Information systems, as well as broad information programs, can be evaluated on a national level. For example, it should be interesting to study the adoption process of new information technologies in Sweden as a part of an evaluation. Criteria such as the speed of adoption, description of users and non-users, early and late adopters, as well as the competition between technologies, should be studied.

Library use

Information use is dependent upon the availability of library resources. From the perspective of libraries, knowledge about user characteristics and behaviour is essential for the planning and administration of library services. In Sweden research libraries have very limited knowledge about who is using the library, what is being used, and the degree of user satisfaction. This can be explained by the fact that libraries generally lack resources to make user surveys and that there is limited knowledge about how to make reliable observations and interpretations of user studies.

Library use is a complex phenomenon that includes several pieces of action: the decision to visit the library, the use of card catalogues, browsing in the shelves, the use of bibliographies and indexes, loans, copy requests, consultations, evaluations, and satisfaction. In order to improve the serviceability of libraries each phase of the process should be studied in detail.

The process of library use can be affected by a large number of factors associated with the users, as well as with the library. The physical availability of library services has been shown to have a major impact on use. The users' own perception of library resources and services is another factor that will affect the use of library facilities. There are also changes in the user population that will affect the demand for library services. Increased specialization of research, part-time students, distance education, and project-oriented education are current examples of changes that create special problems of library access, education, and use.

New technology and communication media in the library environment is another development that will place a greater responsibility on librarians to inform and educate users, as well as themselves. The introduction of computerized information retrieval can be assumed to increase the demand for library service in terms of copy orders and interlibrary loans. Besides these changes, libraries in Sweden have undergone major organizational changes. The university libraries, as well as libraries at smaller colleges, are now integrated in the user organizations and have to compete with other activities when it comes to resource allocation. In this context *systematic studies of library use and development of new services is an essential task for libraries if they shall be able to argue for a better information service*.

Organization of research

Library and information science includes a wide range of research specialities covering various aspects of the communication process. The interdisciplinary character of the field implies that its social and intellectual structure is very complex. Research is performed in several countries, in many different organizations, and by professionals with varying professional compentence and scientific status. Today library and information science is not established as an academic discipline in Sweden.

Considering that Sweden is a small country, one cannot expect Swedish research in library and information science to support more than a limited number of researchers nor to cover all aspects of the field. In the perspective of creating a social basis for research in Sweden, it is also important to coordinate various national projects in order to get an optimum distribution of competence and research. First of all, it seems necessary to establish library and information science in the academic field. The academic status is important, both in order to attract researchers and students to the field, and to stimulate its interaction with other disciplines and fields of study. It is also of importance to rely on those researchers and organizations that are active in the field, if a more formal structure of the field is to be achieved.

There are a lot of academic disciplines where communication problems are studied on a more general basis. Engineering sciences, informatics, computer science, social and behavioral sciences, as well as the linguistic disciplines, have fairly long traditions of research, where information and communication have been the focus of study.

In this paper we have proposed areas of special relevance for the understanding of information needs and uses, the structure of communication systems, and information policy. The various problems that are discussed are mainly of a non-technical nature and should therefore be approached from the perspectives of social and behavioral sciences. Disciplines that have relevance for the research programme are sociology, pedagogy, phychology, economics, philosophy of science, and to some extent linguistics, history, etc. Mass communication research is another academic field of both theoretical and methodological relevance for scientific communication.

In Sweden there is no established Library and information science. There are also very few researchers engaged in studies directed towards the problems of scientific communication. The development of the present activities has involved the establishment of a few information centers and government support for a number of limited studies and development activities. A major part of the development work has been directed towards computerized information services, information retrieval, and computer networks. In the academic field the 'Information Research Group' at the University of Umeå is the only research unit where user research is performed on a more regular basis. In the library field user studies have been performed within some large corporations and sometimes at university libraries. As a consequence of this, there is a lack of systematic and representative studies of information needs and uses, as well as a lack of library research and policy-related studies of scientific communication processes.

Since the number of active researchers studying scientific communication is very small, and since there are only a few ongoing research projects, there is a great need to stimulate research and to develop a plan for the organization of research. To stimulate research and to make the scientific community aware of the problems related to scientific communication should be the starting point for a science policy in this field. This can be achieved using various strategies. To begin with, those researchers already active in the field should be given more stable working conditions in terms of formal positions and long-term projects. There is a great need for attracting graduate students and researchers from different disciplines to define research problems related to scientific communication. The scholarship system already in practice should be given support, but must be supplemented with more active recruitment efforts. Most social scientists and students are unaware of its existence and many might also feel the insecurity of the career opportunities in the field.

Summary

There are a multitude of problems which we need to get a more profound knowledge about. During the last five years the goals and organization of a national information system for scientific and technical communication have been discussed on a policy level. However, the development of a rational information policy should be based on empirical knowledge about the structure and dynamics of the Swedish R&D-system, and the information services and systems available. Such knowledge about user groups and information resources must be followed up by studies of actual information use and communication behaviour. Research projects that can satisfy the need for a systematic and representative description of the scientific and technical information system in Sweden should be given high priority during the next few years.

In a more long-term perspective, basic research about the complexities of communication and information use, as a part of the R&D-process, should be stimulated. In the field of Library and information science there is a lack of theories, which has been pointed out repeatedly by researchers and practitioners. The theoretical shortcomings of the field can be partly attributed to a far too narrow perspective on the communication process. There is a tendency to focus on single media and their present use, without relation to the function of the R&D-system and its actors.

Information use studies need to be integrated with research on R&D-management and science policy. To begin with there is a need for detailed case studies of R&D-work in order to identify the role of information use and barriers restricting the flow of information. Such case studies should also include experiments where new information technologies are tested in natural settings.

The study of information use should not be limited to the exposure to various sources of information. How information is used, selected, and evaluated during the process of problem solving are questions that ought to be asked. Knowledge about these matters can give valuable insights, usable for the adjustment of information services and the management of information flows in R&D-organizations.

Communication problems could also be analyzed on a system level, or on a macro level. R&D-organizations, as well as research areas, are dynamic social systems where communication structures may vary considerably over time. There is a large amount of informal information exchange that complements and sometimes substitutes the formal channels. The structure and development of informal communication is a field of study that is of high priority. Such studies are necessary to understand the relative importance of formal information systems and could also serve as a basis for the assessment of new communication technology.

Considering the wide range of studies that should be made, there is a need for an organization of the field in terms of both research positions and active recruitment strategies. There are many disciplines in social and behavioral sciences where researchers should be made aware of the importance of studying scientific and technical communication. The scientific community at large is one of the main beneficiaries of an efficient communication system, which implies that user research should be performed and evaluated in close relation to the authorities responsible for research policy and higher education.

THE LIBRARIAN'S ROLE AS A MEDIATOR IN FICTIVE LITERATURE

Annelise Mark Pejtersen

> "There are thousands of persons afflicted with a novel reading disease, somewhat akin to the drinking habit..."
>
> J.D. Brown, 1904

> Novel reading, and especially public library novel-reading, has been loaded with the responsibility of inciting people to crime; of suggesting methods of fraud, theft, and even murder; it is charged with tempting senators to neglect their legislative duties, and causing errand-boys to linger on the road, - it is supposed to furnish text-books to every species of moral backsliding, from lying to seduction and bigamy....
>
> *The Library World*, 1902

Introduction

The guidance of a user of fiction asking for a novel with certain topical qualities in a public library calls for a complex professional skill, which it is very difficult to describe in explicit terms. Such a skill is only fully developed in the final library environment since it depends exclusively on intuitive judgement by the librarian. However, to be able to offer librarians a formal teaching and, in particular, to develop suitable library tools, it is necessary to describe the context of the search, i.e. the contents of the book-stock and the aspects of users' needs by a conceptual framework in which the search can be formalized. A classification system for fiction to serve this purpose is described elsewhere, based on an analysis of user-librarian negotiations.[1] Even when the context of the search is formalized by a classification system which is compatible with the dimensions of the users' needs, a librarian will be left with much freedom in his/her guidance of the users. Many books will suit the users' needs which are only rarely defined in all the aspects relevant for the choice of a book. To guide the specific search process, therefore, a librarian needs a performance criterion, a value index to minimize the degrees of freedom for the decisions. Many such criteria are possible: To minimize time spent for a user, to maximize the user's satisfaction etc., but one criterion is particularly important for formal education, since it is stated as one of the basic obligations of professional librarians, namely the criterion of cultural mediation; to support and develop the cultural activities of users.

The performance criterion related to quality aspects of fiction and its role for formulation of proper search strategies is therefore a very important aspect of any classification scheme for fiction and for the design of auxiliary tools.

While a classification system and related search strategies are - to a certain degree at least - objective and rational topics, then the value structures which are necessary as a basis for performance and decision criteria during a search process, are highly subjective and emotional topics. Any choice of value criteria will be open to criticisms.

Legal basis for mediation of values

The use of value criteria to guide the librarian's search to satisfy a user's request is not only necessary to decrease the degrees of freedom in a specific search, it is furthermore typically a request in library laws that librarians should mediate cultural values.

The Danish library law[2] describes the objectives of public libraries as:

> "encourage enlightenment, education and cultural activity"

Criteria for selection of material are stated in the Government directive:

> "As to fiction, readers should have access to the best among available books.... only quality, not religious, political or moral viewpoints must be decisive"

Although book-selection is an implicit way of communicating values, it would be false to jump to the conclusion that the same principle of quality must be administered in the negotiation process with the user, since this might contradict an effort to enlight about cultural activities.

Another directive elaborates on the objective clause quoted above with the demand that

> "materials should make possible independent formation of opinion, general and professional orientation and knowledge acquisition, personal development, artistic and human experience"

Librarians are public employers who are supposed to attend to interests and needs of users as long as these are in accordance with the objectives of libraries. No specific methods, nor definitions of the quoted terms of objectives are given, so the librarian has methodological freedom with an ethical responsibility to pursue these objectives. Traditionally, library laws have expressed the same fundamental view since the beginning of this century, and history shows various perceptions of librarians' roles as intermediaries between the library and its users.

Some librarians felt that it was their obligation to contribute to these objectives by promoting knowledge about facts and history, education, manners and sensibility, the last named by means of the humanities and imaginative literature, and thus make people worthy members of a democratic society[3], and support the task to bring up the population[4].

To others the final goal for the enlightenment of people was the situation, where all classes were knowledgeable to the same extent, which would lead to the dissolution of class distinction in the field of knowledge and information. An attempt was made to formulate a complete description of the perfect active human being that wanted to change his situation, a person which the public library should aim to create and support[5].

The original concept of literature as a means to create rich personalities and to promote individual, worthy manners and a valuable literary taste through knowledge of classical literature and the national cultural heritage, has been dissolved in the present time, due among other factors to increasing specialization and disintegration of traditionally homogeneous classes of society and their educational ideas.

Before the 1960's, this ideal was of predominant interest to librarians but after the economic growth and development, other visions appeared. Social and historical education has been of interest in recent years with the cognition of literature as a specific, historical expression of interests of different classes of society.

The demands of popular genres as the product of 'cultural industry' has convinced some librarians of their duty to fulfil the needs of readers of popular literature with literature of more acceptable values than recreation and amusement. The reason for this is the belief that users' demands for a specific emotional experience from their fiction reading such as suspense, sentimentality, etc., are effected by their circumstances in society and thus fiction can serve as an easy compensation for an unsatisfied need for genuine emotions. The aim is the objective need for improved conditions of life through cognition of reality instead of satisfying spontaneous needs for entertainment.

The discussion began in the fifties but is still of actual interest[5]. A later concept of libraries as democratic institutions implied a limitless right to obtain books of all kinds and stress the respect for the reader's right to make his own judgement[6].

As this brief review demonstrates, goals of literary values have become rather diffuse and difficult to determine unanimously and in generally acceptable terms, compared with the relatively well-established consensus of values in previous times, and it becomes even more difficult to obtain consensus about the application of criteria of literary quality in the librarians' mediation process.

Thus it is necessary to discuss search procedures and the position of criteria of values in the search process independent of decisions of desirable content and goals of the search.

The following is a discussion independent of specific literary values, and to serve this purpose, a rather detailed description of possible applications of different performance criteria of values during a search process is needed.

The mediation process

The librarian is an active and explicit mediator between the library system and the user when he, on behalf of the user, utilizes the available tools and makes selections from the book-stock. The implicit mediating process of book-selection, composition of book-stock, reading lists, displays, etc. is not treated in this paper.

The mediation process is in this context conceived of as a professional task where the librarian has a specific role of search and communication of values, which is recognisable through the librarian's behaviour.

Search is defined as the process during which the librarian with a model of the user's need in his head sets this model up against the library system in a search for the elements, which fit into this model. A search consists of a number of comparisons between the librarian's model and the information he gets from his library surroundings, and it results in a choice. A search consists of several steps and generates potential choices among alternative hypotheses about users' needs, the efficiency of the library system and the character of problem solution that a user will accept. Among a number of relevant alternatives, the librarian will choose the solution which has the greatest value and quality, according to the librarian's subjective performance criterion. The process of search and choice can take many various routes, and the guide to choose a specific one depends on a value criterion.

Communication of values to the user by means of the result of the search is determined by the priorities of the different performance criteria that was used by the librarian during a search.

The choice of value criterion will lead to different forms of mediation, and in the following the communication of literary values as a performance criterion is discussed in relation to mediation categories. (See figure A).

The Role of Librarians in Public Libraries   Figure A

Custodian:	Take care of books; assist users during retrieval; administer auxiliary tools.
Mediator:	Promote cultural values according to library policy and other literary value criteria.
Tools for retrieval:	Are mainly related to topical book content. -The 'aboutness' problem - can be systematized; supported by computers.
Cultural mediation:	Depend on literary value criteria; changing with time, place, social context and personal situation. Tools? Depend on conscious effort of professional librarians.
Retrieval and mediation:	Are functionally related. Many degrees of freedom in meeting the needs of fiction readers. Performance criteria necessary to guide the search. Possible criteria: minimize time; maximize satisfaction; literary value criteria can be effective in guidance if consciously used.

Communication of values is the implicit/explicit transfer of qualities of books to users on the initiative of the librarian. This is a process expressed through verbal communication, and sharply separated from manipulation. Manipulation is defined as the application of methods which efficiently promote the librarian's personal objectives through a conscious misuse of information possessed by the librarian of the library system, which consists of concealing information or making a biased selection of information without informing the user. As an example of practical situations, when the librarian's influence and assistance can be of relevance to users the following can be referred to:

Introduction to new and different literary values through fiction unknown to the user; the imparting of knowledge and understanding of a topic of interest; assistance with personal problems and conflicts and their solution; solution of practical tasks and problems of private or professional origin; and assistance with the improvement of users' individual, social and professional situation.

Concepts of value necessarily control any human act and if fiction is considered as a vital ingredient in the mediating process, it must necessarily be based on a belief in its profound influence on the formation of people's mind and consciousness. In the following, different performance criteria which can be employed by the librarian during the mediation process are suggested. Together with different ways to use the criteria, this leads to different typical mediation processes.

Mediation 1: Passive, neutral mediation. The librarian aims at the exploration of the user's need in order to identify the premises of the user's request in terms of the user's concept of literary value. His performance criterion is the best possible match between the potentials of the library system and his own model of the user's concept of values. The manifestation can be a simple acceptance of the user's request so that the search remains strictly within the frames set up by the initial request of the user, whose idea of value in this case is only implicitly present, or the librarian can actively try to make the frames of the request more significant through a negotiation with the user. When it comes to the librarian's attempts to broaden or narrow the user's criterion of relevance, the operational capabilities of the library system may play a fairly important role.

The user's view of literary value can be of a subjective and private character or in agreement with generally accepted views of value by different social groups. In both cases the user's view is the basis for the librarian's search. The librarian's performance is completely neutral during this search and leaves the choice and the principles of the choice to the user.

One example of experiments with methods for a neutral mediation in libraries has been found. Rubakin believed eagerly in the social functions of public libraries and the user's limitless right and opportunity to obtain books of all kinds. Libraries were determined "first of all by the reader's individual activity and particular aims - everything for the reader, by the reader, in the reader. The librarian's function is to point out books - without putting pressure on him or obstructing his free choice".[7] He realized the lack of necessary tools for creating and sustaining the reader's interest, and developed a statistical method for the evaluation of the relevance of a given text to a given reader, a so called personal 'reading index'. His bibliopsychological method was a mathematical model of a statistical evaluation of readers' experiences in the reading process related to characteristics of given psychological and textual categories. But Rubakin did not achieve his goal to develop general indexes decided for specific social groups at specific times.

Mediation 2: Active neutral mediation. In this situation, the librarian seeks information about the user's need and concepts of literary value as above, but continues with an exploration of the degrees of freedom, while he aims at experiments with alternative value criteria. The librarian's performance criterion is a combination of the user's criterion of fiction value with one or more sets of generally accepted criteria of literary value. This implies an analysis of potential coincidences between users' criteria of value and officially acknowledge criteria of value originating from library tradition, literary history as well as different social groups, such criteria as religious, moral, historical, aesthetic, educational etc.

The manifestation can be a search for books which integrate the desired quality from the user's point of view with a supplementary literary quality. The librarian's performance is value-oriented and he chooses on behalf of the user on the basis of premises originating from users and from institutionalized values.

Mediation 3: Selective objective mediation.

The librarian aims at the development of users' need in relation to his personal, subjective convictions about literary values. The librarian's criterion is in agreement with one or more commonly acknowledged beliefs of literary importance such as the previously mentioned examples in mediation 2. During this search, the performance criterion is to exert influence on users' reading experiences and their concepts of value, and a match between users' beliefs in what is valuable fiction and the criterion adopted by the librarian of what is valuable fiction, is the objective of this mediation.

The manifestation can be a search where the user's criterion of value is implicitly or explicitly known, and is ignored or only partially used as a guide for principle of selection because the librarian shifts between personal preferences and books of user relevance. The librarian's performance is value-oriented and he chooses on behalf of the user on the basis of personal premises or on premises originating both from the borrower's and the librarian's personal values.

If a selective value criterion is narrowly pursued, it is most important that the librarian explains his premises explicitly to the user. If his/her personally selected criterion is concerned with aesthetical and specific literary values it will cause less problems than if the librarian pursues political and moral issues. Previously this distinction between political values and literary schools was rather clear-cut. However, today when clear-cut distinctions between these do not exist, it is necessary to communicate explicitly about criteria of values in order to avoid accusals of propaganda, since the principle of neutrality always has been a well-known feature of the function of public libraries. The distinction between relevant mediation and propaganda is even more difficult to maintain if the personally selected criterion is not generally accepted, but of a subjective and private character, as is the case in the next category.

Mediation 4: Selective subjective mediation.

In this situation, the librarian aims at the communication of personal ideas about valuable fiction of a highly subjective character derived from private experiences. The performance criterion is to exert a personal influence on the user's reading experiences or, perhaps to establish a non-professional communication with the user, possibly of a more private character.

The manifestation will be a search, where the librarian's criterion of value is implicitly or explicitly known, while the user's scheme of literary value is partially or completely ignored. The librarian's performance is value-oriented, and he chooses on behalf of the user on the basis of private premises.

These four categories of mediation represent different degrees of promulgating literary quality: From pure neutrality to pure enthusiasm; from the borrower's personal opinion to the librarian's private conviction. Communication of values in categories 2, 3 and 4 is only relevant if there is a discrepancy between the values of users and of librarians and institutionally accepted values. In category two and three, the librarian communicates generally known literary values, either with a pluralistic or a solistic attitude to values. (See figure B).

Mediation Processes Figure B

1. Passive, neutral	Librarian seeks to meet user's concept of value. No attempt to influence user's criteria.
2. Active, neutral	Librarian seeks to meet user's expressed concept of value, but also explores the user's acceptance of promotion of generally accepted aspects of value left open in the user's formulation.
3. Selective, objective	Librarian attempts selectively to promote one of the generally accepted aspects of value.
4. Selective, subjective	Librarian promotes personal, subjective perception of value.

From a professional point of view, the librarian should master all four types of mediation and possess skills to shift among them.

Mediation and search strategies

The mediation of cultural values by the librarian depends on his use of the freedom to suggest particular books which match the user's needs as expressed by the user or identified during the librarian's search strategy. The possibility of mediation depends on the librarian's use of those aspects of books, which are not specified by the user, to move the user's reading habit towards some higher value criterion. Since users' needs typically are expressed in terms of topical aspects of book contents, the opportunity for a librarian to influence the users' quality aspects is frequently present, or at least, can be considered during the search process.

The opportunity for mediation depends upon the librarian's identification of the freedom he has to influence the user's choice. This depends, of course, upon how specifically the user has identified his own needs, but it also depends upon how well the librarian has explored the user's expressed needs to identify freedom to mediate. In consequence, the mediating opportunity depends mainly upon the search strategy, which is employed by the librarian.

In a previously published study[8] the author has analyzed actual librarian/user conversations to identify and characterise the search strategies used. The typical strategies found are shown in figure C and discussed in the following. There is a natural relation between the various search strategies and the different mediation processes, mentioned in the previous section.

In the bibliographical search strategy the user's expression of needs is the most specific, being in terms of authors and titles. The role of the librarian only is to assist the user to find the books physically or by means of the library tools, and occasionally, to inform the user of the content of specific books.

The Strategies and Roles in Search for Fiction

A1; Bibliographical search with instrumental assistance

The user	The librarian
- explores book-stock	- assists with physical search
- compares book contents and needs	- assists in the use of auxiliary tools
- selects and decides	- explains arrangements and equipments

A2; Bibliographical search with verification assistance

The user	The librarian
- as above	- communicates information on contents of books selected by user

B; Analytical search

The user	The librarian
- communicates information on needs	- explores user's needs
- accepts/rejects proposed books	- compares to contents of book-stock and selects
	- suggests books for user's consideration

C; Empirical search

The user	The librarian
- transmits explicit and implicit signals	- initiates signals by questions and proposals
- accepts/rejects proposed books	- classifies user according to verbal/visual signals and associates to stereotyped set of authors/titles

D; Check routine

The user	The librarian
- requires information on book content	- communicates information on content of selected books
- accepts/rejects	

Figure C.

Literally no freedom is left to the librarian for mediation unless he can initiate a shift in search strategy during the conversation. The empirical search strategy depends on the librarian's classification of users in typical categories based on stereotypical verbal or visual signals from their general appearance together with their initial question formulation, and a subsequent association to more or less standardised sets of books. In this strategy some mediation is possible, but only in an implicit and indirect form depending upon the general quality of the 'standard sets', adopted by the librarian.

An active, neutral mediation or a selective mediation of accepted values can only be perfectly performed during the application of the *analytical search strategy*, which depends upon an exploration of the dimensions of the user's needs, and a search for a book with matching characteristics. Only this strategy will explicitly identify the needs of the user and the degrees of freedom in formulations which can be the basis of mediation of cultural values. This exploration of user-needs can be performed by a discussion explicitly in terms of needs and book contents. Or, the exploration can be done by evaluating the user's opinion about books he/she has previously read, and the content and quality of which are known by the librarian. In both cases, the librarian must have knowledge of the content of books in terms which are compatible with the dimensions and topical expressions of users' needs. A classification system of this kind based on an analysis of user-librarian conversations has been described elsewhere[1]. It classifies books of fiction according to topics, such as subject-matter, frame, author's intention, accessibility.

Such a multifaceted classification can be used as a consistent map in which the librarian can represent the needs expressed by the user. Within the space of this map, which is not specified by the user, the librarian will be free to choose books guided by his mediation criteria, and to explore the value structure of the user, and thus influence the final choice.

This discussion indicates a close relation between the search strategies used and the form of the possible mediation process.

To evaluate how this relation appears in the actual library context, the search strategies[9] have been reviewed together with a previously collected set of conversations to evaluate the cultural mediation taking place. Furthermore, this analysis of actual library conversations also serves to reveal the susceptibility of library users to active mediation.

The mediation role in actual user-librarian negotiations

The following is the results of an analysis of 290 verbal user-librarian negotiations recorded under daily routine conditions in Danish libraries. The purpose is to investigate whether librarians seek to communicate alternative values to the user during the negotiation process and, if yes, how these values are communicated. This study has methodological limitations as to the analysis of the communication, since the complete context of the audio visual communication in the interactions was not registered. This was due not only to the legal, but also to certain practical implications, since the participant users were not informed about the experiment, because it was feared that this knowledge would influence the interaction considerably.

One collection of conversations consisted of a registration of any question about imaginative literature in one public library during a certain period, while the other collection was a selective registration of topical inquiries about imaginative literature collected in various types of public libraries scattered in several communities. Some of these negotiations were not registered by students, but by the librarians themselves immediately after they had taken place. The material in this collection is of an uneven quality, which probably makes the data inconclusive. It is named I in the following tables, while the other collection is named II.

Four categories that constitute the mediation process have been established with the purpose of describing different goals for the mediation, the subsequent role-concept and its manifestation during a search. It is of importance to notice that in the category mediation 1 "passive, neutral mediation", cases have been included in which one can hardly talk about communication of values, but rather of a passive transfer of values, directly within the search. Nevertheless in opposition to a pure search (which is undertaken when a bibliographical strategy is employed) the librarians select books on behalf of the user and thereby communicate values, even if the books proposed belong to the same level of quality as those mentioned as samples of interest to the user. There are various reasons for including these cases in the category, but one obvious reason is to include cases like the coincidence between a user's request for a "valuable" book and the librarian's set of personal or officially accepted literary values. This probably happens in cases where users show a positive interest in serious literature. Regarding the remaining three types of mediating processes, the decisive factor which has been used for the identification of a communication of values is the presence of information in the negotiation process that provides data to classify the titles and authors referred to as belonging to the same quality category or to different categories of quality.

In situations where the only available information consists of the user's inquiries for books such as: Russian literature, modern novels together with the librarian's successive author/title suggestions, the question of premises of values is not answerable because the user's quality level is unknown. This is regardless of whether the reason is the user's refusal to respond to the librarian's attempt to probe his need or the lack of a probing initiative from the librarian. A promulgation of alternative values only takes place when books of higher literary quality or of alternative topical categories are suggested by librarians. It has not been possible to explore what criteria of literary values are preferred by librarians because suggestions of authors and titles cannot be classified exclusively but may possess qualities characteristic of preferences of several literary schools and other institutionalized opinion of values, and because characteristics of what is literarily valuable are hardly ever expressed in the transactions about fiction. Consequently, only pragmatic and well-known, gross categories such as 1. "serious literature" divided into "classics" (before 1945) and "modern literature", 2. "light literature" and 3. "education/information" have been introduced to give an idea of the principles that control the librarian's suggestions of alternatives. The third category expresses a mediation goal of supplementary topical material without any change in the literary level presented by the user. The quality of the fictive works in the interactions is not determined by definitions of these terms, but by their presence in authoritative standard works on literary history of generally accepted values.

Basically the presence of a systematic, explicit mediation will imply the following elements of conversation:

1. an analytical search strategy with its probing of users' needs and criteria of literary value.

2. determination of the freedom in which to experiment with alternative values, as well as users' susceptibility to new proposals.

3. suggestions of titles/authors accompanied by reasons for choice expressed in terms of literary values as well as information about book-contents in terms of users' needs.

4. positive/negative comments to the librarian's suggestions of books with alternative literary or educational values.

In the counting it has been necessary to draw sharp distinctions in order to determine mediation category due to the unclear borders of pieces of communication as well as to the rudimentary elements of explicit communication belonging to a specific category of mediation. These distinctions and the findings are presented in the following discussion.

Users' motivation of mediation

The figures on table 1 show that users are more frequently motivated for a mediation which is also frequently conducted in the conversations registered from selective principles, rather than in the collection with a periodical registration. This is obviously caused by the tendency to collect the interesting cases resulting from requests to select topically on behalf of the user, while the daily routine picture given in the second collection shows a less frequent challenge to execute a search comprising other alternatives than those set up by users. Consequently, the information flow in the negotiations of the latter has a higher percentage of bibliographical data and conditions of the library system (tables 4-7)

However, it is a dominant feature that users seemed very open-minded about librarian's suggestions of books.

It is typical that librarians' usually don't check explicitly the degrees of freedom within which to communicate alternative values during a search, but rather implicitly get an idea of predispositions by reactions to suggestions of more worthwhile titles.

By definition, users are classified as being susceptible to alternative values, when:

1. They initiate an analytical search by leaving the selection of books to the librarian with an inquiry that contains information about desirable book contents in terms of needs, and in the exchange of book suggestions do not put further restrictions on the librarian's search. If on the contrary users show critical reactions by evasion or addition of restrictions to unknown or alternative titles suggested by the librarian, the user is determined to be unsusceptible.
2. They respond positively to the librarian's check of predisposition to a promulgation of alternative values.
3. They accept suggestions of books with alternative values, without having explicitly confirmed a motivation for changes.

It is probable that, in a further examination of their criteria of literary values or in response to information about suggested books in terms of literary values, users would appear less open-minded to active attempts to communicate alternative cultural values.

Finally, the majority of user-librarian interactions follow the goals to execute a search based on users' premises, which naturally will not provoke explicit formulations of restrictions (table 2).

Characteristically, users did not so often explicitly request librarians to assist them in finding alternative responses to their usual needs. Exceptions were seen when users initiated a conversation with a request for guidance for a book within a literary category as for example "modern literature", "German literature", which was motivated by unprofessional desires such as "to move with the times", without much knowledge of the topic. Alternatively, users could directly invite librarians to find something else, something new or different.

Signals of a susceptibility for alternatives are found in situations when users have taken the initiative to ask for a bibliographical search for a specific author/title, which either has not been acquired by the library or is on loan, and they promptly leave the original inquiry with a request for something else, sometimes followed by a refusal to participate in exchanges of information about the need.

Deduction of users' motivations for acceptance of alternatives can be made in cases when users uncritically receive a number of suggestions with various levels of literary quality or when, during the transaction, a user turns out to have read a lot of books with widespread qualities.

These cases with signals of users' motivations for a mediation are most frequently followed up by the librarians, either by suggesting authors at the level indicated by the user, or by alternative proposals.

Librarians' choice of mediation

So far, the discussion has covered the motivation of the user for mediation. Another question is the librarian's motivation for performing an active mediation and his/her choice between different mediation processes.

From table 2, a picture emerges of librarians' conceptions of their mediation role as being predominantly an obligation to fulfil users' needs on their own premises. A mediation on alternative premises is only initiated within institutionalized literary criteria of values, either with or without indications of simultaneously being of personal value to the librarian. A clear picture of a professional attitude emerges with no indications of attempts to communicate private values to users or to manipulate with users.

The concept of the librarian as a neutral mediator seems to have a prevailing acceptance as indicated not only by the large number of instances but also by the fact that mediations 2 and 3 often are initiated when a problem with the availability of requested material arises. In the communication of literary categories, serious fiction plays the most important role in librarians' choices of alternatives.

An implicit mediation process is also observed when librarians, probably through unconscious associative leaps, supported by a mind's eye picture of the book stock on the shelves, utilize their personal knowledge of imaginative literature, which frequently, due to their level of education, is on a more advanced level than that expressed by users who turn to the librarian for help.

The mediation of light fiction as alternatives to users' inquiries is never fulfilled, but it appears when an empirical search is initiated with a stereotypical set of titles of light fiction, and the user's subsequent rejection leads to a search for more worthwhile literature. The previous identification of a high frequency use of empirical search strategies[8] is not in principle inconsistant with nor will it impede a communication of alternative values. Although an empirical search from a logical viewpoint is not the ideal choice of strategy, it has the advantage of placing very little demand on the librarian's memory and skills of negotiation. If it is to be employed in a search that aims at the development of users' concepts of cultural value, then the librarian must be able to operate with other stereotypical author/title sets of serious literature, perhaps even with sets of titles selected on the basis of historical criteria of value accepted by different movements of literary criticism and schools. Finally, the skill to generate sets of titles of various qualities from a relevant choice of performance criteria is needed.

When librarians take the initiative to explicitly propose a search involving alternatives to users' inquiries, this initiative is often motivated by a lack of relevant titles within the subject area of interest to the user, either because relevant titles are popular and on loan or because their quality is too low to be acquired by the library. The same pattern is seen here for both user and librarian in that both were unanimous in explicit demonstration of motivations for mediation, when conditions of the library system and its policies put limits on the satisfaction of users' spontaneously expressed needs.

Luckham[10] stated from controlled interviews, that readers were often unspecific. They did not identify or relate a specific theme to their books but had only generalized subject interests. He found that if topical interests, or genre, were of importance, choices of specific titles were frequently a consequence of external influences. This points at a possibility for librarians to influence users' choices of authors, provided that users do have confidence in librarians' ability to advise them. Recent research[11] reveals that users are not very expectant of help from library staff.

Spiller's investigation of how fiction readers choose their material in public libraries found the same generality and openmindedness in fiction users' approaches to the fiction stock. Most respondents had a small list of favourite authors, long since exhausted, and were anxious to discover new names. The majority carried something between five and ten names in their heads and resorted to browsing if books by their familiar authors were not available. In the present study, user-librarian interactions indicated the same pattern of knowledge of a small number of authors together with a lack of memory of what had been read earlier, unless this information could be recollected through recognition of titles in the librarian's questions about previous readings. Furthermore, a fixed standard set of authorships are employed both by user and librarian as a means of quick communication of categories of needs and quality levels.

Information flow in the negotiation process

A more detailed analysis of the information flow in user-librarian negotiations is concerned with the kind of information that librarians seek from users in searches of imaginative literature, as well as the type of information that is transferred from librarians to users in relation to negotiations of authors/titles.

Table 5 provides a clear impression of the title/author communication as a prevailing implicit method to get information about users' needs, while the more straightforward interview of users' desires in terms of book contents is less frequently employed. This is consistent with the general low priority given analytical searches and the widespread preference for empirical searches, where librarians, in succession, check the user's knowledge of individual titles from the standard repertoire.

It is also seen that the initiative for a mediation is usually taken without an initial check of motivation and this confirms the impression of rather implicit or spontaneous initiatives.

This is in agreement with table 5, where a detailed counting of questions about book aspects illustrates the weight placed upon aspects other than author's intention of cognition/information. This information is considered to be of great importance for a mediation of cultural values, because information about a user's demand for qualities of cognition/information will give a hint of the premises for their concepts of literary value. No direct exchange of information about specific literary value criteria affiliated to literary schools were present.

The exchange of information with relation to titles of relevance in the communication (table 6) shows the emphasis on book contents as well as on bibliographical data and data about the library system. Again information about author's intention of cognition/information is absent, while an interestingly high number of specific data on time and place aspects of books is found (table 7).

Probably this is related to fundamental principles of the capacity of human memory; information is both easier to learn and remember if it can be located in time and place[12]. Time and place categories are easily recognised and do not involve semantic problems, which makes them convenient in a brief negotiation situation.

There is a slight tendency to negotiate about the negative aspects of books which are not wanted by users, instead of an effort to formulate positive criteria for a search. It is stated by Alexander that while negative criteria of value can be explicated, positive value criteria can only be determined through experiments[13]. In this case a possible explication of the librarian's frequent probing of users' needs through a call for previous readings of specific authors/titles together with subsequent suggestions of book-titles can be considered as a method of experiments with confrontations, intended to provoke users' quality judgements.

Difficulties with formulations of positive expectances from reading seem to be a common phenomenon. Luckham found that people were more specific in their description of the actual effect of book reading than in their characterization of the expectances from books beforehand, especially in connection with interviews about their "deep inner experiences" with books. It turned out that people who expected entertainment from selected books, often changed to an attitude of acceptance with gaining experiences of human and social conditions, and were satisfied with this reading effect.

Luckham's data pointed at a gap between respondents' formulation of criteria of selection and their actual experience and acceptance of book contents. This underlines the idea of freedom to experiment with cultural values.

Presumably users' questions to librarians will contain another type of information than those employed to express the actual positive experiences of preferred books. This indicates a significant field for librarians' potential exploration of users' positive/negative expectances and experiences reading effects as carried out through interviews to uncover emotional experience and cognition/information derived from specific titles, given that users will be motivated for this kind of conversation.

The empirical data discussed in table 1-3 should be regarded as preliminary results due to the implicit character of communication of values, but could be further tested through a statistical analysis of questionnaires answered by users and librarians.

Conclusion

The need for professionally skilled librarians in advising and selecting books for users of fiction has recently been recognised by Spiller[11]. A considerable number of users (46%) had chosen their novels after browsing through the fiction stock and a strong traditional element was apparent in the restricted set of authors which were referred to again and again.

Furthermore, the majority of readers (88%) found that blurbs were the most important factor when they selected books. That these are uncertain sources of information was demonstrated by the fact that the return of unread books was due most frequently to a dislike of these books.

Readers are not accustomed to asking librarians for assistance, and Spiller finds that the librarian's attitude to fiction service is partly responsible for this. This leads to reflections on the concept of librarianship and the advocated skill of the profession.

Dervin's study of the conceptual factors of the identity of professional librarians demonstrated the placing at a high priority on the manipulation of library tools and the prestige connected with the challenges of complex tasks of information search, while an understanding of users' individual and subjective needs was not considered as part of the professional skills required in librarianship, and thus has had a low priority[14].

A study of organizational factors in the library profession emphasizes ambiguity of roles and needs for instructions about expectations to the professional staff[15].

Finally, communication skills of librarians have been subject to many studies[16, 17, 18]. They generally conclude that the library staff as a whole is hesistant in taking up the interpretation of the librarian as a professional expert in communication with users. Even if these conclusions are of a very general character and discussed independently of various contexts, they point at a priority of the custodian role of the profession, while the librarian's role as an intermediary between users and the fiction stock demands a broader concept of the professional role. The needs of fiction users in opposition to the requirements of non-fiction users are generally not related to a specific subject-matter, but to several different features, and thus a number of books with common features will be able to satisfy the user equally well. If consciously used, literary value as a criterion for selection among many relevant books can be effective in the guidance of the librarian's search.

Only to some degree, value aspects can be represented by a systematic classification system. The dimension called "Author's intention" in this author's scheme will reflect the author's potential, serious intention of communicating values, education or information. The "Accessibility dimension" can represent the specific 'literary' value in terms of reference to language and literary form.

However, in general, value judgements are changing with time, social and educational background. In addition, ideas of literary values are today most inhomogenous and should not be included in classification schemes and development of auxiliary tools, except perhaps as additional information on ideological affiliations or affiliations with literary history and schools within well-known standard categories as Catholicism, Classics, Romanticism, etc.

A classification system representing book contents alone will be a reasonable invariant and can form the basis for the design of tools to support the librarian's search strategies. The guidance of the search by relevant value criteria, however, depends on a professional librarian's currently updated perception of values and mediation obligations. It is, therefore, important to consider the development in the roles of librarians due to the advent of effective tools utilizing computers. These might ultimately change the librarian's role from being custodians of tools and book-stock to active mediators. Computers can overtake the neutral manipulation of auxiliary tools, while instructions for judgement of values as a performance criterion in various, individual situations are extremely difficult to design. The conscious process of the mediation of values is a unique skill which a professional librarian should possess in the form of knowledge to place authorships within a given context of generally accepted values. Users' criteria of topical interests are explicitly expressed by the categories of the AMP classification scheme for fiction. Literary values of books are only present implicitly in terms of categories and the description of book contents. The explicit criteria of literary values, however, existing within different literary schools and periods, which provide given books and authorships with specific characteristics of value, must be part of the librarian's professional knowledge, as well as the ability to utilize these value criteria in relation to a topical classification of books. It is also part of the librarian's skill to know the function of different criteria of literary value during a search process and utilize them in a search and mediation process in relation to a classification scheme. Furthermore, knowledge of the various mediation processes and their application to different, more or less well-formulated users is necessary.

The librarian's professionally thorough reading of each individual book and total absorption of its contents will be of less importance (due to a potential assistance from a classification scheme), while wide reading of fiction of different qualities from different periods of literary history and criticism, of society and cultural values etc. will be most important to a mediating role.

If it is considered advantageous that the mediating role of the librarian becomes more explicit and systematic, it will be necessary to support such a change by a precise formulation of rules for the use of the various and changing value criteria as guides for search of fiction. Such rules should be integrated in the instructions for the application of any classification system and other retrieval tools.

Frequency of Mediation Table 1

	156 conversations I	134 conversations II	Total 290 conversations
Users motivated	104	40	144
Users not motivated	19	9	28
Mediation initiated x)	24	14	38

x) only instances of active mediation counted (Type 2 and 3)

Frequency of category of mediation process Table 2

	156 conversations I	134 conversations II	Total 290 conversations
1. Passive, neutral mediation	105	38	143
2. Active, neutral mediation	20	12	32
3. Selective, objective mediation	4	2	6
4. Selective, subjective mediation	0	0	0

Frequency of communication of literary categories Table 3

Type of mediation	156 conversations I	134 conversations II	Total 290 conversations
Passive, neutral			
Serious fiction	60	23	83
Classics	26	14	40
Modern	34	9	43
Light fiction	75	21	96
Active, neutral & Selective, objective			
Serious fiction	27	16	43
Classics	12	8	20
Modern	15	8	23
Education/information	7	2	9
(Light fiction)	4	1	5

Frequency of librarian's questions Table 4

Category of question	156 conversations I	134 conversations II	Total 290 conversations
Bibliographical data and library administration	14	63	77
Users' previous readings	56	23	79
open questions	37	9	46
closed topical questions	19	24	33
Users' previous readings of specific authors/titles	75	129	204
Users' needs in terms of book contents (see table 5)	46	17	63
Check for motivation for alternatives	7	11	18
Other			17

Categories of contents of questions about users' needs Table 5

Terms of users' needs	156 conversations I	134 conversations II	Total 290 conversations
Subject-matter	13	2	15
Frame: Time, place, social/professional setting	9	5	14
Author's intention			
Emotional experience	10	4	14
Cognition/information	0	0	0
Accessibility	4	7	11
Readability	2	0	2
Literary form	2	7	9
Other			14

Frequency of librarians' information transfer in relation to book suggestions Table 6

Category of information	156 conversations I	134 conversations II	Total 290 conversations
Bibliographical data, library system and administration	97	199	296
Book contents (see table 7)	125	114	239
Librarian's personal knowledge and recommendation	14	22	36
Information about authorship	7	8	15
References to other media: T.V. etc.	3	8	11

Frequency of category of information about book contents Table 7

Terms of users' needs	156 conversations I	134 conversations II	Total 290 conversations
Subject-matter	32	23	55
Frame: Time, place, social/professional setting	28	34	62
Authors intention	19	23	42
Emotional experience	18	23	41
Cognition/information	1	0	1
Accessibility	12	5	17
Readability	12	5	
Literary form	0	0	
Other	0	0	0

References

1. Pejtersen, Annelise Mark. Design of a classification scheme for fiction based on an analysis of actual user-librarian communication, and use of the scheme for control of librarians' search strategies. In: Theory and Application of Information Research. Ed. Harbo, Ole and Kajberg, Leif, 1980.

2. Lov om folkebiblioteker. 1964. Bekendtgørelse om folkebiblioteker. 1965.

3. Kirkegaard, Preben. Folkeopdragelse. Bogens Verden, 1940.

4. Jacobi, Carl. Digtningen og Folkeopdragelsen. Bogens Verden, 1938.

5. Hovman, Ole. Den dårlige smag. Bibliotekaren, 1956.

6. Agger, Jens Peter. Folkebibliotek eller elitebibliotek. Biblioteksdebat. 1969-70.

7. Simsova, S. Nicholas Rubakhin and Bibliopsychologi. 1968.

8. Pejtersen, Annelise Mark. Investigation of search strategies in fiction based on an analysis of 134 user-librarian conversations. In: IRFIS 3. Proceedings of the 3rd International Research Forum in Information Science. Ed. Henriksen, Tor. 1979.

9. Pejtersen, Annelise Mark. Låner-bibliotekarsamtaler. I og II. Danmarks Biblioteksskole, 1974, 1976.

10. Luckham, Bryan. Bookreaders' Expectations and Experiences. University of Manchester. 1980.

11. Spiller, David. The provision of fiction for public libraries. Journal of Librarianship, 1980.

12. Miller, George A. Psychologi and information. American Documentation. 1968.

13. Alexander, Christopher. Notes on the Synthesis of Form, 1964.

14. Dervin, Brenda. Information Needs of Urban Residents, I-III, 1977.

15. Stead, Bette Ann, Scamell, Rich. W. A study of the relationship of role conflict, the need for role clarity and job satisfaction for professional librarians. The Library Quarterly, 1980.

16. Black, Sandra M. Personality - librarians as communicators. Canadian Library Journal, 1981.

17. Luckham, Bryan. Increasing Inequality. A Century of Regression in Public Library Service. University of Manchester. 1981.

18. Barnes, Marion. Staff-User Interaction in Public Libraries. Crus, 1980.

PROBLEM SOLVING AND INFORMATION RETRIEVAL. A CASE.

Povl Timmerman

We have recently finished a study where we have given reference questions to librarians and users and studied their search behaviour by means of a thinking aloud technique. We have considered (manual) information retrieval as a problem solving task, and this has several implications. Firstly we have considered cognition or thinking in the light of a so called frame theory put forward in cognitive science. The essence of the frame theory is this: When one encounters a new situation or makes a substantial change in one's view of a situation, one selects from memory a structure called a frame. A frame is a data-structure for representing a stereotyped situation like being at a conference or going to a child's birthday party (Minsky, 1977).

In our study we designed a reference question like this:

> "The weather forecast offered by the local telephone administration gives the degreeday figure for the day before. Find a definition of degreeday figure."

We expected the librarians to produce a meteorological frame and search books in meteorology. And this happened. Our move was to make sure that no book in this class would provide an answer. So we created a problem for the librarians. And we wanted to study in which way the librarians changed frames in order to find an answer providing tool.

We now turn to a second implication of the problem solving theory. It is usual to separate problem solving strategies into two distinct categories:

algorithmic methods

heuristic methods

An algorithm is a step by step method, where the sequence of steps is well known. If each step is performed correctly, the method is guaranteed to lead to a solution, if given enough time. If no algorithm is known or the known algorithm is impractical one must resort to heuristics, which are rules of thumb, strategies, or shortcuts aiding discovery of a solution. Heuristics are not guaranteed to find optimal (or any) solution, but often produce satisfactory solutions in a reasonable amount of time.

An objective in our study was to discover such heuristics when trying to solve the problem we had created. At the time of analysis we had no ideas of the nature of such heuristics. In the meantime Marcia I. Bates (Bates, 1979) published two articles with some 46 formulated heuristics or tactics as she calls them. To give some examples:

S4 Parallel	to make the search formulation broader by including synonyms or otherwise conceptually parallel terms.
T5 Trace	to examine information already found in the search in order to find additional terms to be used in furthering the search.

In our study we found:

1. Librarians are likely to use only the terms in the question
2. Librarians are likely to go directly from topic to answer providing tools
3. Only reference librarians are attempting to collect additional terms from answer providing tools which themselves are negative

Our participants seem not to use some of the tactics which in some cases could solve the search situation.

It is our opinion that this situation could be somewhat remedied by introducing the concepts of heuristics and presenting exercises where specific tactics could be useful. This requires that we make exercises where we have a fair amount of control over the search paths the students are likely to follow. I shall exemplify how such exercises can be designed.

Design of exercises

We know that the terms used in a written reference question is likely to be used as search terms. We can, therefore, select such terms as to block direct access to answers.

We have as an example designed a question around the following topic:

> "Find a law regulation, which gives a map of sailing regulations around the Dan-field."

Now, this question contains some shared knowledge, since any librarian and almost any student can relate the Dan-field to one of the Danish oil wells in the North Sea.

Law material is in a law collection which has an alphabetic index and a systematic index. The systematic index has a group of 'sailing regulation'. But, for some reasons the document is not entered here, but in a group named 'Underground'. The index terms are safety zone, underground, continental shelf, but not sailing or navigational terms. The year of issue is not given, so the retrospective search covers 6-8 years.

From the point of view of formal training in reference work, the law material collection is the correct answer providing tool. But the viewpoint has been that of the producers of documents and their way of organizing and indexing the material.

The viewpoint of users of the particular document will lead us to

"Fisherman's handbook", which has the indexwords safety zone and Dan-field, but the amount of law material is so small that a page-by-page search will easily give the result.

The heuristic working here is to go from document producer to document user. The corresponding more general tactic by Bates says:

F4 cut when selecting among several ways to search a given query to choose the option that cuts out or eliminates the largest part of the search domain at once.

Teaching situation

We can now from the topic design different reference questions and reasonably well control the difficulties in the search.

Formulations like these can be given to different groups of students:

Find a map of the safety zone around the Dan-field.

Find law material indicating the safety zone around the Dan-field.

Find law material indicating sailing regulations around the Dan-field.

Find law material indicating sailing regulations around the Danish oil wells in the North Sea.

In this way we can study the effect of chosen terms on the search paths.

The setting we have been using is to have 4 students in a group. Two students are asked to make the search and talk about the problems encountered in a collaborative problem solving. Two students are asked to be observers and write down a diary of the whole process.

In this way we get some form of a written protocol which forms a basis for discussing search problems and identifying heuristics acting in problem situations.

The educational aspects are covered theoretically by Aluri (Aluri, 1981), who deals with Ausubel's cognitive theories in educational psychology. Aluri favours problem solving tasks as best suited in library-use instruction to train and measure meaningful learning.

Conclusions

Theories and ideas put forward in cognitive science are models of reality. They can be used to create and analyse a reference situation in training situations.

From a point of view of research, student sessions can act as pilot studies and indicate problem spaces i.e. all possible paths leading to solutions.

From a point of view of education, exposure to problem situations and definitions of tactics will have a lasting effect and help the librarians (and the users) to broaden searches and cope with particular indexing problems.

It seems important in natural language searching to pay attention to a specific training of librarians. Our study indicates that things do not just settle by the use of retrieval tools.

References

1. Minsky, Marvin. Frame-system theory. In: Thinking. Readings in Cognitive Science. Johnson-Laird, P.N. and P.C. Watson (eds.). Cambridge, Cambridge University Press, 1977.
2. Bates, M.J. Information search tactics. Journal of American Society for Information Science, July 1979.
3. Bates, M.J. Idea tactics. Journal of American Society for Information Science, Sept. 1979.
4. Aluri, Rao. Application of learning theories to library-use instruction. Libri, vol. 31, no. 2, Aug. 1981.

PITFALLS OF USER ORIENTED RESEARCH, AND SOME NEGLECTED AREAS

J.M. Brittain

Introduction

Drawing upon the results of user studies of the last 30 years, carried out for the most part in an attempt to assess the information and document requirements of professionals, I shall summarise their general characteristics and objectives from a critical point of view, and then consider aspects of the study of information needs and information seeking behaviour that have been relatively neglected. I shall not consider the problems associated with the assessment of the information requirements of members of the general public, because in this area the methods and approaches that have been developed to study the information needs of professionals are quite inappropriate.

In dealing with neglected areas I shall touch upon social, political, and philosophical issues surrounding the concept of information need, information use, and so on.

Characteristics and objectives of user study research

As an area of serious study we can trace work back to the period immediately following the second World War. At the 1947 Royal Society Conference two seminal papers were presented by Bernal and Urquhart, which provided a great stimulus to the study of the information, document, and library requirements of scientists. For many years user research was on a small scale, carried out by librarians (less often by scientists themselves), usually with the objective of improving information and library services at a local level. The use of questionnaire, and less frequently the interview technique, soon became the predominant methods of user study research.

For at least the first decade the objectives of these studies were modest. Basically, scientists were asked questions about the documents and the information they used in carrying out their professional activities. Most users (it is interesting to note that potential users were of no interest for many years!) had a reasonable working knowledge of the journal literature of their discipline and at least an elementary knowledge of ways of obtaining documents. Scientists were asked questions only about documents, literature and library services that they had some knowledge of, and some experience in using. They were not asked hypothetical questions of the nature "If you had known about service X would it have been of use, made a difference to your research, affected the findings, etc.?" This type of question was a feature of the more ambitious and complex user studies of the late 1960s, and of the 1970s.

In the early days the question of the application of the results was not usually raised, mainly because few attempts were made to apply the results outside the context in which they were obtained. Also, few attempts were made to generalise the results from one group of scientists to scientists in general. In many cases, the applications of the results were self-evident. For example, if a scientist said he had difficulty in obtaining certain documents, attempts could be made by the local library to improve the supply situation.

The early studies were also uninterested in theories or rules of user information seeking behaviour. They did not attempt to provide a theoretical structure of information requirements; indeed, little attention was paid to the comparison of results of one user study with another, or to the accumulation of the results of user studies to provide a body of knowledge about information requirements in general.

The first decade of user studies lasted for about ten years, that is into the late 1950s, by which time user studies had become much more ambitious and wide-ranging, and more complex in terms of methods of assessment, treatment and presentation of data. Researchers were no longer content to ask professionals about information and document services they were familiar with, but asked about potential information needs. By this time a large number of studies already existed and it was inevitable that researchers would begin to try and compare the results of one study with another, in order to make sense of seemingly disparate findings from different environments and groups of users. Also problems were raised about the application of the results of user studies to the improvement of existing services and the design of new ones.

Although these topics have received a good deal of attention during the last twenty years, many have not been satisfactorily resolved and continue to be raised from time to time. In this paper I shall explore some of the reasons why user studies have run into a number of seemingly intractable problems. There has been much dissatisfaction with the progress made in the study of information requirements. Since the first issue of the Annual Review of Information Science and Technology (ARIST), which for a number of years contained an overview of user studies in its first chapter, many reasons have been put forward to explain the unsatisfactory nature of many user studies. Since the early 1970s the chapter on user studies has appeared less regularly, but nevertheless still serves as a good focus for some of the discontent, criticisms, and problems that have concerned those involved in user study research. I shall enumerate some of these areas of criticism and discontent, with the aim of throwing light upon the ways in which improvements can be made in the study of information requirements. However, first we must consider briefly the major methods and approaches used to assess information needs.

Major areas of study

As we have already noted, the early studies were limited in scope and for the most part asked scientists direct questions about their needs for information, documents, and library services. But much research and writing about user information requirements has been interested in other aspects of information seeking behaviour. We can identify at least four major areas of study.

1. Direct enquiry of users, either by questionnaire or interview, about the documents, information and library tools and services they actually use; the extent of use, the purpose of the use, and with what success and effect.

2. Assessment of the attitudes of users towards library and information services. For example, users may be asked if they like a particular bibliographical tool or information service in its present form, what changes they would like to see, and so forth.

3. Experimental information services and their evaluation. A selected group of users are presented with an experimental information service (e.g. a new form of abstracting indexing service, an SDI service, etc.), and subsequently, after using the service, users are asked, either by questionnaire or interview, about the usefulness of the service, their attitude towards the service, and suggestions for improvements.

4. Direct observation of users, usually in operational situations. For example, researchers may observe library users at the catalogue, in the process of using bibliographical tools or computerized and on-line information retrieval services. There have been relatively few such studies, perhaps because there are relatively few aspects of information need and information seeking behaviour that can be observed directly in this way, without questioning users.

Questionnaires and interviews have been used widely throughout the community of library and information users, beginning with the early studies of professional scientists, extending to applied scientists and practitioners, social scientists and social science practitioners in the 1960s, and to members of the general public in the 1970s. One of the problems of present day user studies is that the basic methods have gone largely unchanged, although the objectives of user studies have changed appreciably over the years.

Many other methods have been tried infrequently (e.g. diary keeping, job analysis), but have met with relatively little success.

There have been many suggestions for the use of methods derived from social science research, but little systematic development of new methods.

Problems of traditional user study research

It is now widely acknowledged that there are a number of seemingly intractable problems involved in user study research. These problems have now been discussed for at least the last twenty years, although there is still no satisfactory resolution of the majority of them. In order to understand the reasons for this unsatisfactory situation it is necessary to consider aspects of user study research in the context of the nature of science and social science information and knowledge. Few enquiries at this level have been reported, although I suggest that it will be a most fruitful area of study in the next decade. I now mention briefly some of the major problems of user studies of the last two decades, giving examples drawn from my own studies of the social sciences, and also from other studies in science, applied science, and applied social science.

1. Application of results of user studies to the design of new information and library services and the improvement of existing ones. This is the area that most interests practising librarians and information service managers. In the early years the specific information and document requirements of a handful of scientists were studied in a particular institution. The results were easy to apply. The information and document needs of particular scientists were either being met or were not being met; if the latter the local library could take steps necessary to meet the requirements. Less often, in the early studies, scientists would suggest new information services, and appropriate action could be taken. As user studies became wider in scope, the information requirements that were assessed were often of a more general nature. A typical statement from latter day user studies is of the order "In getting to know about new references x per cent of the respondents said they used personal recommendations, y per cent bibliographical tools, z per cent browsing". This type of finding increases our general understanding of the communication

patterns of professionals, and could be described as a study of the sociology of communications, especially where the communications are related to institutional, personal, and demographic variables. The fact that the results are of relatively little use to an operating library or information service has gone unchallenged. Even librarians themselves have carried out this type of study without realising that the results are very difficult, and in some cases impossible to apply, at least without a considerable amount of further development work. There are now hundreds of such studies in nearly all fields of science, applied science, social science, administration, and the humanities. Undoubtly we have a better picture of the average user of libraries and information services than we did twenty-five years ago. An important general finding is that users are not systematic and rational in their seeking after information and knowledge. Indeed, one of the findings that is common to nearly all the studies is that users do not like bibliographical tools for information and document retrieval, do not use them rationally and systematically, and much prefer (often by a factor of 10 : 1) informal to formal channels of information. Few practising librarians have acted upon these findings. Obviously the implications would have been too unpleasant to face - in some institutions the logical conclusion would have been to disband the formal communication and library service altogether!

2. Methods and methodology. Although many methods have been tried in the assessment of information requirements, questionnaires and interviews have predominated. These methods have proved reasonably successful when the objective has been limited to an assessment of the document and information requirements of scientists who are relatively regular users of libraries and documents, and who have a reasonable familiarity with library tools and libraries. They have been less successful in the assessment of potential information requirements, and the requirements of professionals who are not bibliographically oriented and whose information needs, although varied, are not met by libraries, library services and bibliographical tools. The majority of user studies have been library and document oriented, rather than consumer oriented, or work oriented.

It is only relatively recently that user studies have attempted to assess in detail work situations and work problems, and then to relate, or deduce from these findings, the information needed to successfully and efficiently carry out work tasks, and solve scientific, social, economic, political and practical problems.

The questionnaire and interview are used extensively in many areas of the social sciences; they are the basic methods of survey research. The methods have been very little changed when applied to library research. They are in the tradition (as indeed are the other less frequently used methods such as diary keeping, indirect observation, participant observation) of quantitative, empirical research in the social sciences. Although there has been a good deal of discussion about the inappropriateness of these methods for library research, especially in assessments of potential need and the information needs of users who are not library and bibliographically oriented, there have been few developments. There is still dissatisfaction with these methods. In the last decade an entirely different approach has been tried by way of qualitative assessment of information services and of the information gathering activities of users. However, these studies are few in number, and have not been found to be entirely satisfactory (see, for example, Harris, 1981; Hounsell and Winn, 1981; Oldman, 1981; Wilson, 1981).

3. Documents v. information/data. The majority of studies have been concerned with the need for and methods of finding out about, and obtaining, documents; few user studies have been concerned with the requirements for data or information as such. One of the reasons for this is undoubtedly the fact that the early studies of scientists, mainly researchers - who are usually fairly knowledgeable about libraries, documents and so forth - paid a good deal of attention to the requirements of scientists for keeping up-to-date with their own specialities. It was quite obvious that scientists did this, at least in part, by reading journal articles, research reports, and so forth. Questionnaires and interviews are appropriate for finding out in general terms about the ways in which scien-

tists keep up-to-date with their own field, and get to know about other fields as necessary. These methods are not so appropriate for the assessment of specific needs for information and data. Research scientists have always had a need for such; increasingly social scientists, management scientists, and administrators have a need for data and information, as opposed to a need for documents only. User studies have not kept pace with the changing patterns of use.

4. Use made of documents. Although user studies have concentrated upon the methods by which users get to know of bibliographical references and the ways in which they retrieve documents, very little attention has been paid to what happens to documents when they arrive upon a user's desk. How long do users spend reading documents of different types; are documents typically required for one fact or idea, or for their general information; how long do users need to keep documents; could the information in them be supplied in more efficient and effective forms? These, and many other questions about the use made of documents, have been almost totally neglected in library research. This is perhaps because most library and information services are evaluated in terms of the number of documents supplied, and to a lesser extent their relevance. Obviously, a service that supplies ten relevant documents is better than one that supplies fifty documents of doubtful relevance, other factors being equal. If more attention was given to the relationship between documents, data, and information on the one hand, and the solution of scientific, social and economic problems on the other, and also the relationship of information to decision making, no doubt library researchers would pay more attention to what use is made of documents.

5. Information overload. The belief of most practising librarians has been that information per se is good; the more the better. There has been little or no concern about users' ability to digest and make use of information, or about the optimum flow of information in terms of work tasks and users' ability to process information. We should not be too critical of library researchers on this front. These problems are

formidable indeed, and social scientists, including computer scientists and psychologists, have themselves not been particularly successful in researching into human information processing.

6. Differences between major user groups. As user studies proliferated in the late 1950s and more so in the 1960s attempts were made to compare the results of one study with another and also to provide general statements about the information requirements of major categories of users, and contrast major groups one with another. At the time of INFROSS (Investigations into Information Requirements of the Social Sciences) attempts were made to compare the results from social science user studies one with another, and also to compare major social science users with science users (Bath University, 1971; Line, 1971). It proved to be very difficult (see, for example, Skelton, 1973) because of an absence of standardisation of methods of assessment, data presentation, and analysis. It was seen that the findings one study from another were in little agreement, albeit the same major group of users was being investigated. It has generally been concluded that social scientists do not differ fundamentally from scientists in terms of their information requirements. Yet the absence of substantial differences between major categories of users may be illusory, and hidden by the inappropriate methods of assessment, or the inability of existing studies to tease out some fundamental and important differences. These matters have been discussed in some detail elsewhere (see Brittain, 1979).

7. The development of a body of knowledge about user requirements. One of the major problems is that the results of many hundreds of user studies have not accumulated to form a body of knowledge about the information needs of users, from which conclusions can be drawn about the way existing information services can be improved, and about the design and development of new ones. The results of the studies of the past twenty-five years undoubtedly increase our general understanding of the way in which users, especially professionals, use (and very often do not use) libraries and information services.

But many of the results are of more interest perhaps to the sociologist of science, or even the philosopher of science, than they are to the average practising librarian or information scientist.

Neglected aspects

I would like to mention four neglected aspects of information need and information use.

1. Communications not involving documents. It has been variously estimated that between 50 and 80 per cent of the communications of the average researcher and academic take place through informal channels of communication (e.g. face-to-face meetings at conferences, telephone calls, seminars, discussion groups, etc.) Since it has often been assumed that the bread and butter of the researcher and the academic is the formal communication channel this comes as some surprise, or at least should do, but has rarely been taken very seriously by librarians. Outside academia the amount of informal communication is even greater, and many professionals in social work, and other areas of applied social science, administration, and teaching rely extensively upon informal communications, and make relatively little use of the vast number of documents that are potentially available to them, and which librarians have traditionally thought are of potential value and interest. In many situations the work of practitioners revolves around clients, local information, local gossip and hearsay, as well as the more traditional informal communications such as conferences and meetings. Most professionals have some needs for documents, but usually very precise and limited in nature to directories, legal regulations relating to a particular occupation, and so forth. Most user studies have paid far more attention to the formal channels of communication than to the informal channels. This is not surprising since many user studies have been undertaken by librarians, in the context of libraries and information services. This emphasis was acceptable when user studies were carried out in institutions of research and higher education, but as user studies were extended into environments, and especially into areas

where professionals do not have easy and ready access to documents and library and information services, the concentration upon the formal communications system became less and less satisfactory. Practical knowledge and practical know-how are not usually transmitted through documents.

2. Use made of information/documents. User studies have concentrated upon the demands users make for documents, information, and library services. Very few have paid attention to what happens to documents once they reach users. What do users do with documents, how do they read them, what do users do with the data and information supplied by library services? Use made of documents is related to the virtually untouched area, at least for the information and library profession, of the cognitive processes involved in reading, problem solving, and making use of data and information in the solution of problems and in the creation of new knowledge. Surely it is the intellectual activities - the problems that are solved and the knowledge that is created - that are the raison d'être of library and information services? No-one would pretend that merely supplying documents is a useful and valuable activity in itself. And yet, although the use made of the ideas and information contained in documents is of paramount importance, and should be related to the amount of information that flows, and the way in which it is supplied, information scientists and librarians have for the most part ignored it.

3. Relevance and reinforcement. Although information scientists and librarians have paid a good deal of attention to the concept of the relevance of retrieved documents, they have ignored for the most part the concept of reinforcement as developed in psychology. A few academic users - theoretical and laboratory scientists - may indeed find nearly any type of information reinforcing, but the vast majority of users require only highly relevant information for specific work tasks. The amount of information supplied, the context in which it is supplied, and the flow of information over time are variables which affect the level of reinforcement. These contingencies of supply and demand have been neglected. There is scope for much research and development at this level. Different users

and different categories of users are likely to function at different levels in terms of the supply characteristics. Documents and information services need to be 'tuned' to estimates of the probability of user interest and user success in solving scientific and social problems. A start was made many years ago in constructing user profiles for SDI services (and this has long since been a routine part of many services), but much more could be done to ensure that users receive information that is of current interest, value, and applicability.

4. Finally, in considering neglected aspects of user studies, we must take into account the work involved in producing 'maps' of the literature by bibliometric methods. Citation analysis provides a valuable tool in the construction of 'maps' of the literature, which indirectly relate to the structure of knowledge in scientific and social scientific fields. This work is still in its infancy and 'maps' of the literature have been produced only in a few areas so far. Practising librarians have been very slow to see the potential of this work.

The future?

Most library and information services make the assumption that the more bibliographical references, the more documents and the more information supplied the better the service. Library and information services have in the past been evaluated on a rather crude measure of quantity rather than quality. If a librarian supplies users with many documents then they are seen to be, and no doubt themselves feel they are, doing a good job. It is not surprising therefore that the vast majority of user studies have tried to tease out all possible information requirements, including unstated and potential needs. It is obviously a way of creating a greater demand for services of a traditional nature. In this paper we have considered some of the methodological problems of user studies that have attempted to measure more than the current demands for documents by scientists who are library and bibliographically knowledgeable, but more importantly, we have considered some of the almost totally neglected aspects of information gathering seeking and use. If these were to be researched, and if the results of research in the future were to be acted upon, we would see the development of a very different and new generation of library and information services, in which the flow of information was much more finely tuned to work tasks and problem solving.

The traditional user study will continue, albeit on a reduced scale, perhaps at a local level, and with some demonstrable results, but the challenge for the future resides elsewhere - that is, in the matching of library and information services to the structure of knowledge in particular disciplines, and in the development of a new generation of information services that supply highly relevant information, to the right people, at the right time, and with the express purpose of solving social and scientific problems and creating new knowledge. Perhaps this new area of study should be called by a name other than user studies, or information needs studies? An appropriate description of this field of activity may be human information processing in the context of the solution of social and scientific problems.

Conclusions

1. If the objectives of user studies over the last 30 years had remained relatively unambitious - at the level of simple and direct questions to professionals, mainly scientists, as was the case for the first decade - the methods of study, the analysis of the data, and the application of the results would have presented few problems. However, the objectives of user studies soon became much extended, but without concomitant changes and developments in techniques of measurement, and in the absence of standardised methods of measurement, data analysis, presentation, reporting and application.

2. The number of user studies increased greatly in the 1960s and 1970s. This resulted in an enormous quantity of data about the information gathering, seeking, and using behaviour of a large number of scientists, applied scientists, social scientists, practitioners, administrators, government officials, schoolteachers, educationalists, and finally, members of the general public. There were no generally acceptable theoretical

guidelines to make sense of this huge mass of data; no guidance in the design of studies, the standardisation of methods, and so forth. It was soon seen to be impossible to compare the results of one study with another, and to draw general conclusions about the information requirements of any particular group of users, in a way that the results could be made use of to improve services or design new ones.

3. Although the results from the large-scale studies provide insights into the communication activities of users, the major studies were usually very one-sided (i.e. from a document and library point of view), and pay little attention to the vastly important area of informal communication.

4. In the future the small-scale study at a local level is likely to continue, but it will have little to contribute to the understanding of communication activities in general.

5. On a broader front, issues and problems outside the immediate area that has become designated as user studies are likely to play an increasingly important part in determining the direction of future studies of information requirements. Aspects of communication that may demand our attention in the future include: information overload and the negative effects of information; informal channels of information; the need for greater selection and filtering in the supply of information to the majority of users (who are not academics and researchers) in order to provide limited amounts of highly relevant information and data, at the right time, in the right place, in a readily acceptable and usable form; a movement away from services that supply bibliographical references only; a movement away from document based services towards services that are truly data and information retrieval services; a greater emphasis on the use made of documents, especially the value of information in the context of work tasks and problem solving.

6. The tremendous investment that now exists in traditional library and information services suggests that major changes in document and information delivery will be slow.

7. Information scientists must be prepared to change their traditional ways of evaluating and assessing services; in particular there must be a movement away from assessment in terms of quantity of items supplied, towards one in which quality (in terms of value and use) is given greater emphasis.

8. The most radical change in our orientation to users would involve a movement away from the concept of the information requirements of users to one in which more attention was paid to the information requirements of work tasks; this in turn would direct emphasis upon the use made of information in the solution of scientific and social problems.

References

1. BATH UNIVERSITY. Investigation into Information Requirements of the Social Sciences (INFROSS): Information Requirements of Researchers in the Social Sciences, 2 vols. Bath University Library, May 1971. (OSTI Report 5096).

2. BRITTAIN, J.M. Information services and the structure of knowledge in the social sciences. International Social Science Journal, 1979, 31(4), 711-728.

3. HARRIS, C. The travelling workshops experiment: an attempt at 'illuminative evaluation'. Social Science Information Studies, 1981, 1(4), 247-253.

4. HOUNSELL, D. and WINN, V. Qualitative approaches to the study of information problems. Social Science Information Studies, 1981, 1(4), 203-207.

5. LINE, M.B. The information uses and needs of social scientists: an overview of INFROSS. Aslib Proceedings, 1971, 23 (8), 412-434.

6. OLDMAN, D. Sources of antagonism to 'qualitative' research. Social Science Information Studies, 1981, 1(4), 231-240.

7. SKELTON, B. Scientists and social scientists as information users: a comparison of results of science user studies with the investigation into information requirements of the social sciences. Journal of Librarianship, 1973, 5(2), 138-156.

8. WILSON, T.D. A case study in qualitative research? Social Science Information Studies, 1981, 1(4), 241-246.

REPORT ON WORKSHOP 1: MODELS AND THEORIES OF USERS' NEEDS

Chairpersons: Kathleen Bivins (Monday)
Nicholas Belkin (Monday-Wednesday)
(Minutes or edited discussion notes)

The notes below follow our discussions during the three sessions from Monday to Wednesday. We had several temporal directions for the work-group and many conflicting and unsettled views which, as I believe, to a great extent were due to misunderstandings and conceptual problems. I believe that these are obvious in the notes. They are not complete and reflect my conception of the discussions.

1. Monday-session, September 14

The principal agenda was set to discuss the following:

- perspective on models and theories
- naming of problems for research
- reasons for doing information need (IN) studies
- how do people manage without - as they obviously do - the Traditional Services?
- how does information use relate to productivity?

(This list of themes was not actually followed any further).

The question how to design new information services was posed. Studying the use of today's services as a possible answer was seriously questioned. That could reveal some current problems in the current context but would not reveal what features are desirable for the new services.

Then an attempt was committed (excuse me for the funny expression, that's not by accident) to list factors influencing information seeking (IS), either positively (inducing) or negatively (prohibiting). It turned out, however, that each proposed factor of either type could be turned upside down and seen to have effects of the other type. So the work-group decided not to list more factors.

It was pointed out that in doing his work, the work actualizes information and INs in the person's mind. It was questioned, how needs, in general, are related to INs and how these, again, to the IS behavior. Much talk about this led us to ask, are there any assumptions taken for granted in the IN research. Some were listed:

- the rationality of users
- supply orientation - there is information to look for in IN situations
- information is a value
- there are information needs
- expertise (user's expertise vs.mediator's expertise)
- truthfulness (of service demands and of information)
- verbalization of needs is possible; a common context of users and mediators so that a mutual understanding may develop
- consensus and will for cooperation throughout the chain from information suppliers to mediators and end users

A new agenda was set for the Tuesday session:

- discuss a general model of user studies

- discuss the contribution of IRFIS 4 presentations to the group work

- how the results of user studies can/should be used

- the methodology of user studies

We decided to have T.D. Wilson's model (J. of Documenation 37 (1981)1, March 1981, pp. 3-15) as a frame of reference.

2. Tuesday-session, September 15

T.D. Wilson's model was discussed for a while. Among others, the following points were raised:

- the model resembles much that of Paisley's from the sixties (cf. C.A. Cuadra (ed), Annual Review of Information Science and Technology, vol 3 (1969), pp. 1-30).

- how should unconscious needs be dealt with in the model?

- are the physiological needs really important for information needs research; shouldn't we be more concerned with the cognitive aspects?

- how should time be dealt with in the model?

It was noted that time and the environment, consisting of the working environment, the work itself and persons, influence IS behavior.

Some other models were also drawn up (they are not described here). We noted that there are many different models for INs or IS behavior and that they can be characterized as procedural models, modelling the steps involved in fulfilling INs, and declarative (or environmental) models, modelling the factors influencing either INs or IS behavior (or both). It should also be possible to find a meta model for the multitude of models.

The work-group decided hereafter to concentrate on the starting point of IS. What kind of a model explains this? We noted the inducing and prohibiting factors touched upon (and rejected) in the Monday-session would be relevant here. Also relevant might be the means-ends-analysis (viewing the decision situation involved as a gap between a person's goals and resources).

Then we had a debate on whether we should take information needs for granted or start to explain how the needs turn up. We had no consensus here.

We listed the agenda for the Wednesday-session as follows:

- prepare the report of the work-group
- list physical, social and cognitive factors relevant to the work-group theme
- list issues for research

3. Wednesday-session, September 16

(Due to the arrangements and some misunderstandings this session lasted only some 15 minutes with several work-group members missing).

The following points were laid down for the report:

- What is observable is the IS behavior.
- Based on this we postulate there is an IN.
- We focused on the starting point of IS where the need is transformed to IS. Here we ran into severe conceptual problems and quarrelled about whether we should take the needs for granted and work forwards from them or start with the observed IS behavior and work backwards to the transformation point.
- We noted that both concepts, information and need,

are context dependent and multidimensional, having many, possibly in part contradictory, meanings. We had nc consensus on these.

- We wondered what happened if we had such a consensus - what advantages would be gained? The first one is that then we would have no need to discuss the matter any longer.

- We believe that the external factors (e.g. available channels) influence the IS behavior much more than the INs.

- We have conflicting views: a) its appropriate to study INs by studying IS behavior and b) this is not appropriate.

Kalervo Järvelin

REPORT ON WORKSHOP 2: INFLUENCES FROM OTHER DISCIPLINES: BORROWING AND LOANING

Chairperson: Povl Timmerman

1. The philosophical basis of library/information science

The workshop members, invited to contemplate the diagram (Fig 1) provided by the Chairperson and to comment critically on it, became quickly conscious of the diversity of their academic disciplines and experience. We decided to confess, each to the others, what we had done in the information world and what we were now interested in. After those revelations we took refuge in philosophy in the hope of finding a common basis for our further discussions.

The 'three-world' model of Karl Popper offered a formulation which withstood criticism from some members. Popper recognises:

World I - the physical world of which our bodies are part;

World II - the subjective mental world, the world of 'mental states' or the totality of human minds;

World III - the objective world of human knowledge enshrined in the artefacts now scattered over the Earth. These artefacts include not only all the documents stored in our libraries but all physical forms which have been created by mankind.

It is now unhappily possible to conceive of a nuclear calamity which could generate enough radiation to wipe out all human life on Earth and yet leave our artefacts intact. When the radiation had sufficiently declined, it would be possible for other forms of intelligent life to visit Earth and note the artefacts. When

Lending service

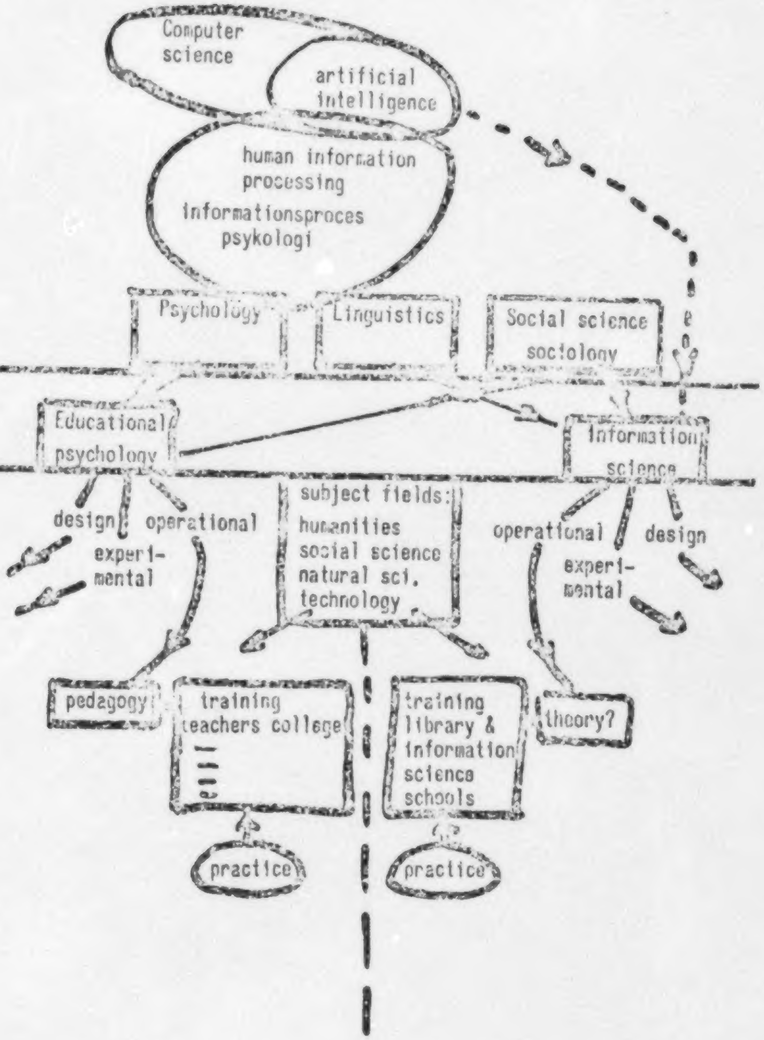


Figure 1

interpreted, they would reveal the story of human life on Earth and what we learned about the physical world and about ourselves.

However, the 'objective knowledge' enshrined in our artefacts, even in our documents, is not immediately accessible to those who seek it; there is much redundancy and even contradiction. But the task of library/information science could be described as that of exploring World III and organising it for more effective use in World II.

2. Information science: a natural or a human science?

The complaint was strongly made that because of the 'institutionalizing' of the humanities and sciences, it was impossible to unify the worldpictures of humanists and scientists and therefore impossible to teach information science in the same way to both humanists and scientists. To present information science as a component of the 'science of science' had its difficulties too. And the suggestion that it might be regarded as an aspect of political science also failed to win support. However, it was argued that there was no reason why humanists should not be taught in one way and scientists in another. There is plenty of work for both to do and the particular skills of both can be fully exercised.

3. What should information science borrow from other disciplines?

The diagram was again brought to our notice. Forceful objections to the idea that information science should borrow from other disciplines were voiced. If it did, it would become merely a 'rag-bag of bits and pieces'. Instead, it was argued, information science must adopt a 'do-it-yourself' policy, finding its own way to theoretical coherence. It was admitted, however, that it might well be found after integration had been achieved that information science had borrowed from elsewhere. But then the borrowed items would also have been transformed.

Gradually we appreciated that the diagram could be interpreted in other ways; it shows how things presently seem to be but does not necessarily present a program.

4. General systems theory and the science of information

One of our members, a systems theorist, explained that, from his point of view, information science could be subsumed under general systems theory. The three key elements of systems theory were matter, energy and information. It was the introduction of information into dynamic physical systems which differentiated systems theory from physics.

This argument won some sympathy though it was pointed out that information science is concerned with systems in which information is wholly cognitive and definable as that which changes a state of knowledge. Systems theory seemed not to admit knowledge.

It was therefore agreed that it was desirable to distinguish the study of information processes in purely physical systems from what is now known as information science by calling it the science of information.

5. Conclusions

We were all surprised that the Chairperson's purely objective diagram of the present state of information science and its relationship to other disciplines could be so systematically misinterpreted in so many different ways. It certainly provoked a lively discussion. Was it a diagram or a work of art?

Bertram C. Brookes

REPORT ON WORKSHOP 3: PRACTICAL RESEARCH APPLICATIONS AND PROJECTS

Chairperson: Tor Henriksen

According to the organizers we were supposed to relate discussions as much as possible to the papers given during the three days. We first of all shared each other's impressions of the different papers given. Then we tried to understand the main ideas and had some discussions on them, and finally we asked ourselves what kind of practical consequences or applications we can have from the ideas presented in these papers. We tried to follow the same schema for all the papers but obviously they were all very different, so we couldn't always follow it.

The first day we started with Nick Belkin's paper, 'Models of discourse and information retrieval system design' and I think we all found that it was a very interesting model, and we had the pleasure of having the author there to give further explanations, but as the research itself was not started yet, we found out that we had to wait for the results, and he told us that it would be in a couple of years, so then we will come back to possible practical applications of that one.

Considering the next one, 'Language as a barrier to communication in the social sciences' by Ralph Adam, we discussed first the reason for this barrier because we agreed that there was a kind of barrier there, and we ended up with the conclusion that at least people within information science and mediators of information in general should be very much aware of the special terminology situation in the social sciences. There was no immediate application project to build on that, personally I added that people within the social sciences might be encouraged to do more terminology work and, as I already said, information science people should be aware of the special problems.

Then we turned to 'Action research and users' needs' by David Streatfield. David did not have the problem of application because it was applied already, so from a workshop viewpoint it was a very good paper. It was applied in the method, so we discussed the paper instead and had a very interesting discussion about that. We also found out that, maybe, it is easier to get money for such projects because it is so easy to show results because they are inherent in the project. There were some remarks on the term 'action research'. I remember Ole Harbo said that in his paradigm this was not action research and at least we recognized that if that was action research many of us had been doing action research without knowing it. So there is a little terminological problem here.

Then we turned to 'The education of the end-user in computerized information retrieval' by Nancy Fjällbrant. We had a very long discussion on user education, and we disagreed a little on how much time or effort should be spent in educating the end-user. Some obviously thought it was too much, others not enough, and some people said that it is alright to do some education of the end-user for the sake of motivation, so they will be aware of these sources, but several members of the group thought that information retrieval is too difficult to be left to end-users.

Then we came to 'The mediation and use of medical information' by Tuula Laaksovirta from Finland, and we looked at the results and we had a kind of confirmation from the author of the paper that she had stopped with the results and had not been willing to go very much further with conclusions, so we suggested some conclusions. I remember one was that the pharmaceutical industry perhaps had too much impact in the information work, but that was not one of the conclusions drawn in the paper, so we continued to discuss possible conclusions then.

The next morning, we listened to Kathleen Bivins 'Methods in user-oriented research: A critical review of 'Thinking Aloud". We naturally focussed on the method because it was the kind of research that proved things we more or less felt that we knew already, but now it was perhaps still more time than before that users have a tendency to use rather broad terms when formulating their search. But it showed us another possible application area of this 'thinking aloud' technique.

Then we had 'The concept of entailment for the purpose of mediation' by Marta Grabowska. That paper was not easy to understand for all of us because we were not all used to symbolic logic, so we spent some time during the workshop to try to understand and we also asked the author to explain to us some of the main ideas of the paper another time. Afterwards we discussed whether this way of dealing with documents, translation of documents, paraphrases of documents, and abstracts of documents was useful. There was a disagreement. There were two parties: One thought what it was not really interesting to do this, but others thought that it was really interesting, and we reached no conclusion. But I think the majority thought that it was a new and interesting way of attacking categorization of documents.

Then we had a look at this data base with government reports from Sweden, and we agreed that it was a kind of free text search system with access to a limited part of the document, in particular the descriptions of the diagrams and tables and of the contents pages, and we discussed the experimental project related to what is normally believed to be suitable in free text search. We found that the first condition to have it, let us say, more than 90% in one language, was fulfilled. The second one was that it covered a limited number of years. The third one we doubted a little because we thought that it was quite spread over a wide variety of subjects, but I heard afterwards from Björn Tell, that that was not the case. The terminology of this field was quite clear so most conditions for having a successful free text search are there, and we look forward to see the evaluation of this data base before we can conclude and tell if this is applicable in other fields.

Then we discussed 'The wasteful use of empirical statistical information in information research' by B.C. Brookes. And here was some confusion and that was perhaps my fault because we were using the word information in different meanings. At least we agreed that the main point, if we have understood the author right, the main point is that we should be careful in using the result of statistical investigations in situations where there was one single user. There was no disagreement on that, but I had a personal observation there, that I repeated in the group, that maybe we are using the word information in too many meanings, so sometimes we do not understand each other. Here statistical information is used where other persons would have used statistical data instead. So, I think, one day we shall have a definition of how we are allowed to use the word information, because it is used, and also during this conference, in many meanings.

Then we listened to a research programme for Sweden, 'Research on the needs and uses of scientific and technical information'. It was a very detailed programme and it was a paper that was very different from the other ones, so it was difficult to relate it to possible applications because first we must see how it will be applied in Sweden and what results there will be, before other countries copy it, so we stopped at that level.

This morning we discussed Annelise Mark Pejtersen's 'Theory and practice of the librarian's role as a mediator in fictive literature'. It was a very long paper, and we had to choose one single topic of greatest possible generality to discuss it in the very limited amount of time that we had got. So we chose to discuss her model 'from passive objectivity to engaged subjectivity' and found out that that could be useful also outside fictive literature - as Stephan Schwarz said already just after the presentation that he would like to be quite far down on this scale but he refused to be placed at a certain point of it. So we had a certain discussion there, but we should not be afraid of having librarians who are engaged and show their own personality. On the other hand, it is a profession and professionals should, to a certain amount, hide their personal opinions when they are doing service to users. One of our conclusions was that many of the findings seemed to be applicable in information science in general, also for non-fictive literatures, so it would be worth while trying it.

Then we moved to Povl Timmerman's 'Problem solving in information retrieval' and we discussed his division between algorithmic and heuristic problem solving. One of the participants, I think it was Hans Holm, said that this is a very sharp division and in reality it is not like that. You mix the two, you start with a plan, and then something happens and then you have to use the trial-and-error method which we in scientific conferences call heuristic problem solving methods. On the other hand Ole Harbo said that that does not matter, it is anyhow good to get categories to work from and even if you mix them in practical work it is useful to know what we mean and he was rather positive towards this way of dividing things.

Finally, Michael Brittain's paper was discussed in the context of the next conference.

As a whole there were not many papers that we could just take and look at the result and say, "Oh, that's fine! Let us do it in practical work". So perhaps we have to have a 5th, and a 6th and a 7th IRFIS to reach that point, but, after all, for some of the papers we thought there are things to follow up here immediately, for example in curricula or in practical work.

Tor Henriksen

LIST OF PARTICIPANTS

Ralph Adam

The City University,
Northampton Square,
London EC1V OHB
England

✓ Nicholas Belkin

Centre for Information Science
The City University,
Northampton Square,
London EC1V OHB
England

✓ Kathleen Bivins

The SCANNET Foundation
c/o Medicindata
Box 33031
S-400 33 Göteborg
Sweden

Odd Brisner

Norsk senter for informatikk
Forskningsveien 1
Oslo 3
Norway

✓ Michael Brittain

Dept. Library & Information Studies
Loughborough University
Loughborough, Leics.
England

✓ Bertram C Brookes

64 Abbotts Gardens,
London N2
England

Svend Bruhns

Royal School of Librarianship
Aalborgafdelningen
Langagervej 4
DK-9220 Aalborg Øst
Denmark

Mats Cavallin

Göteborgs universitetsbibliotek
Box 5096
S-402 22 Göteborg
Sweden

Göte Edström

Högskolan i Borås
Box 55067
S-500 05 Borås
Sweden

Nils-Bertil Faxén

Högskolan i Borås
Box 55067
S-500 05 Borås
Sweden

Nancy Fjällbrant

Chalmers University of Technology
S-412 96 Göteborg
Sweden

Ingegerd Friberg

(Högskolan i Borås)
Stampgatan 58 B
S-411 01 Göteborg
Sweden

Hans-Ove Frid

Swedish University of Agricultural Sciences
S-750 07 Uppsala
Sweden

Mariam Ginman

Strandst. 18
SF-00330 Helsinki 33
Finland

Marta M Grabowska

Instytut Bibliotekoznawstwa
i Informacji Naukowej
Uniwersytet Warszawski
Ul. Nowy Swiat 69
00-046 Warszawa
Poland

Charlotte Hanner

Högskolan i Borås
Box 55067
S-500 05 Borås
Sweden

Ole Harbo

Handelshøjskolens Bibliotek
Julius Thomsens Plads 10
DK-1925 København V
Denmark

Tor Henriksen

Statens Bibliotekshögskole
Dalenenggatan 26
Oslo 5
Norway

Inger Heyerdahl-Jensen

Royal School of Librarianship
Aalborgafdelningen
Langagervej 4
DK-9220 Aalborg Øst
Denmark

Roland Hjerppe

Delegation for Scientific
and Technical Information
P.O. Box 43033
S-100 72 Stockholm
Sweden

Hans I Holm

Haessle AB
S-431 83 Mölndal
Sweden

✓ Lars Höglund

Inforsk, Sociologiska institutionen
Umeå universitet
S-901 87 Umeå
Sweden

Peter Ingwersen

Royal School of Librarianship
Birketinget 6
DK-2306 S København
Denmark

Carl Gustav Johannsen

Royal School of Librarianship
Aalborgafdelningen
Langagervej 4
DK-9200 Aalborg Øst
Denmark

Kalervo Järvelin

University of Tampere
Dept. of Library and Information Science
Box 607
SF-33100 Tampere 10
Finland

Søren Kaae

Danish Center of Library
Service
Telegrafvej 5
DK-2750 Ballerup
Denmark

Leif Kajberg

Royal School of Librarianship
Birketinget 6
DK-2306 S København
Denmark

Merja Karivalo

Helsinki University of Technology
Otakaari 1
SF-02150 Espoo 15
Finland

Maj Klasson

Linköpings Universitetsbibliotek
S-581 83 Linköping
Sweden

Tuula H Laaksovirta

The Academy of Finland
The Medical Research Council
Rahamestarink 12
SF-00520 Helsinki 52
Finland

Chagan Lalloo

Högskolan i Borås
Inst bibliotekshögskolan
Box 55067
S-500 05 Borås
Sweden

Agneta Lantz

Linköpings Universitetsbibliotek
S-581 83 Linköping
Sweden

Marit Lauten

Statens Bibliotekshögskole
Dalenenggatan 26
Oslo 5
Norway

Larserik Lönn

Högskolan i Borås
Inst bibliotekshögskolan
Box 55067
S-500 05 Borås
Sweden

Kari Marklund

Linköpings universitetsbibliotek
S-581 83 Linköping
Sweden

Annelise Mark Pejtersen

Royal School of Librarianship
Birketinget 6
DK-2306 S København
Denmark

Ursula Nielsen

Lilla Brogatan 41
S-502 30 Borås
Sweden

Rolf-Allan Norrmosse

Högskolan i Borås
Box 55067
S-500 05 Borås
Sweden

Lena Olsson

Linköpings Universitetsbibliotek
S-581 83 Linköping
Sweden

Gerd Paul

Wissenschaftszentrum Berlin
Bibliothek und Dokumentation
Platz der Luftbrücke 2-4
D-1000 Berlin 42
BRD

✓ Olle Persson

Inforsk, Sociologiska inst.
Umeå universitet
S-901 87 Umeå
Sweden

Bertil och Kerstin Quentzer

Högskolan i Borås
Inst bibliotekshögskolan
Box 55067
S-500 05 Borås
Sweden

Rein Saarnak

Bibliotekstjänst AB
Box 1706
S-221 01 Lund
Sweden

Stephan Schwarz

Royal Institute of Technology
S-100 44 Stockholm
Sweden

Lars Seldén

Högskolan i Borås
Inst bibliotekshögskolan
Box 55067
S-500 05 Borås
Sweden

✓ David Streatfield

National Foundation for Educational Research
The Mere, Upton Park,
Slough, Berkshire
England

Ralph Strömfelt

Sveriges Mekanförbund
Box 5506
S-114 85 Stockholm
Sweden

Ritva Sundquist

Helsinki University of Technology Library
Otaniementie 9
SF-02150 Espoo 15
Finland

Björn Tell

Lunds universitetsbibliotek
Box 1010
S-221 03 Lund
Sweden

✓Povl Timmerman

Royal School of Librarianship
Birketinget 6
DK-2306 S København
Denmark

Pertti Vakkari

University of Tampere
Dept. of Library and Information Science
Box 607
SF-33100 Tampere 10
Finland

Birgit Wanting

Royal School of Librarianship
Aalborgafdelningen
Langagervej 4
DK-9230 Aalborg Øst
Denmark

Elsa Werngren

DFI
Box 43033
S-100 72 Stockholm
Sweden

Irene Wormell

Lunds universitet
Box 1010
S-221 03 Lund
Sweden